DOCUMENT RESUME

ED 031 222

JC 690 262

By-Cohen, Arthur M.; Brawer, Florence B.
Measuring Faculty Performance.
American Association of Junior Colleges, Washington, D.C.; California Univ., Los Angeles. ERIC Clearinghouse for Junior Coll. Information.
Pub Date 69
Note-90p.
Available from-American Association of Junior Colleges, 1315 Sixteenth Street, N.W., Washington, D.C. 20036 ($2.00)
EDRS Price MF-$0.50 HC-$4.60
Descriptors-*Effective Teaching, Evaluation, *Evaluation Criteria, Instruction, *Junior Colleges, *Performance Criteria, Personnel Evaluation, *Teacher Evaluation, Teachers

This report focuses on the techniques, objectives, and problems associated with teacher and teaching evaluation. The first section of this 2-part monograph was devoted to the discussion and appraisal of methods currently employed. Topics considered included: a review of the standard techniques of supervisor, colleague, student, and self evaluation; a discussion of the problems inherent in these methods such as rater bias, ambiguity of purpose, and lack of definite criteria; an overview of research attempting to relate teacher personality with teacher effectiveness; and an outline of a program incorporated into the junior college teacher preparation program at UCLA that aimed at predicting the success of new teachers by the use of various personality dimensions. A case for changing the purposes, methods, and criteria of faculty assessment was presented in part two. It was suggested that faculty evaluation, as a tool to improve instruction, must relate to instruction as a discipline with the focus placed on the effects of instruction, an approach that may result in the development of team teaching techniques and evaluation among instructors on the basis of teaching effectiveness alone. The problems in specifying criteria for assessing teacher effectiveness, a rational for using student achievement of learning objectives as the main criteria of teacher effect, suggested designs for assessing instructors, and a scheme for supervising instruction were presented. (MB)

ED031222

NCR - do not film - info. reproduced on
[illegible]

MEASURING FACULTY PERFORMANCE

By Arthur M. Cohen and Florence B. Brawer

JC 690 262

ERIC Clearinghouse for Junior College Information / American Association of Junior Colleges

The ERIC Clearinghouse for Junior College Information operates under contract with the United States Department of Health, Education and Welfare as a joint project of the University of California, Los Angeles, Graduate School of Education and the University Library. This monograph was prepared pursuant to that contract. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent or show official Office of Education position or policy.

Price: $2.00

American Association of Junior Colleges
1315 Sixteenth Street, N.W.
Washington, D.C. 20036

Printed in U.S.A.

MONOGRAPH SERIES

Salvage, Redirection, or Custody? Remedial Education in the Community Junior College. By John E. Roueche. 1968. $2.00.

Junior College Institutional Research: The State of the Art. By John E. Roueche and John R. Boggs. 1968. $2.00.

Personality Characteristics of College and University Faculty: Implications for the Community College. By Florence B. Brawer. 1968. $2.00.

Measuring Faculty Performance. By Arthur M. Cohen and Florence B. Brawer. 1969. $2.00.

FOREWORD

The community junior college enrolls a great variety of students. Many of these are in the process of identifying their personal, educational, and vocational goals. Students may range from antagonistic to enthusiastic in their attitudes toward the learning process. Some are self-propelled; others require persuasion and a skillful cultivation of tentative interests. Consequently, basic to achievement of the objectives of junior colleges are two functions performed by professional staff—teaching and counseling. The teacher is an essential participant in the learning experience of junior college students. So essential is the teacher that the quality of his work will in large part determine whether the junior college will fulfill its particular mission in education. It is important, therefore, that faculty effect be assessed and thereby improved.

Cohen and Brawer maintain that an institution dedicated to teaching—as the junior college is—should study instructors, students, and the learning process. They propose that student gain toward specific learning objectives be recognized as the ultimate criterion in assessing effects of teachers and teaching situations. They suggest acceptance of "causing learning" as a definition of teaching and maintain that such learning can be appraised in an objective fashion. If this definition is accepted, the criterion for each evaluation must become a demonstration of student learning which may be presumed to result from the efforts of the teacher in question. The reader will recognize the difficulty in isolating effect from a particular learning situation; nevertheless, the approach has a good deal more to commend it than those commonly used which often fail to differentiate between the teacher as a social being and the effects of his teaching.

It should be obvious—but unfortunately it is often forgotten—that assessment of student gain is futile unless instructors specify clearly what they are trying to teach and what measures they intend to use to assess learning. In this connection, one of the benefits of programmed instruction is a built-in insistence upon precise objectives or outcomes.

The authors raise the question of objectives for evaluation. Why are we interested in measuring instructors in junior colleges? They charge that the purposes are nebulous; as typically conducted, faculty evaluation cannot be seen as a way to improve instruction.

They suggest several ways to improve instruction: the teaching profession might well begin to police itself in order to counter external judgment. All persons working in the institution should not be expected to be thoroughly competent in all facets of instruction, but rather the institution should be staffed by a core of people who collectively, but not necessarily individually, display excellence in all matters related to teaching. What is needed is to discover *who* can teach *whom*.

At this time, a great deal of discussion has centered around the need for more than ten thousand new junior college teachers each year. Several models of preparation have been proposed and can be found in various universities and colleges. Cohen and Brawer assert that the junior colleges themselves must take a larger responsibility for preparing their own instructors; university-based teacher preparation programs are not proving adequate. There can be no question about the need for junior colleges to marshall their own expertise and to devise rational and logical ways to measure faculty performance in order that instruction can be improved. This monograph is a provocative and timely aid in that direction.

Edmund J. Gleazer

Executive Director
American Association of Jr. Colleges
Washington, D.C.

PREFACE

The Educational Resources Information Center (ERIC) is a United States Office of Education endeavor. The ERIC Clearinghouse for Junior College Information, one of eighteen in the ERIC system, was established in June 1966. Arthur M. Cohen, assistant professor of Higher Education at U.C.L.A., is principal investigator and director of the project; Lorraine Mathies, head of the Education Psychology Library, is coinvestigator.

The Clearinghouse collects, indexes, and abstracts documents containing information relative to all phases of junior college operations —students, staff, plant, curriculums, and organization. Its particular acquisitions emphasis is on research studies produced by junior colleges and on publications reporting results of research concerning junior colleges. In addition to its indexing-abstracting function, the Clearinghouse is charged with information analysis and synthesis. Accordingly, interpretive documents are produced in the monographs, topical papers, and other substantive reports.

This monograph, the fourth in the Clearinghouse/AAJC series, examines an issue that directly affects all junior college administrators and faculty members. Every college either has—or has consciously rejected—a scheme of faculty evaluation. Some plans operate well; many more exist only because no one has thought of anything better. In the monograph, merits and deficiencies of various plans for assessing teachers are discussed and alternatives suggested.

Also reported here are results of original research conducted in the U.C.L.A. junior college teaching internship program. The program's operation has been fully described in *Focus On Learning: Preparing Teachers for the Two Year College*, Occasional Report #11, Junior College Leadership Program, available from the U.C.L.A. Students' Store.

Both authors are affiliated with the U.C.L.A. Graduate School of Education as well as with the Clearinghouse. Arthur M. Cohen directs the junior college teacher preparation programs and teaches courses on the junior college; Florence B. Brawer is a research associate in the department.

The American Association of Junior Colleges has been generous in its support of the Clearinghouse's efforts. Our special thanks to members of the Association for their generosity in providing funds to publish this report and to the United States Office of Education for making possible its production.

Arthur M. Cohen
ERIC Clearinghouse for Junior College Information

CONTENTS

Page

Foreword I
Preface III
Introduction VII

PART I

Chapter I. **The Media of Faculty Measurement** 3
Chapter II. **Evaluation Through Personality Variables** 17
Chapter III. **Ratings of Junior College Teaching Interns** 28

PART II

Chapter IV. **A Critique of Current Practices** 51
Chapter V. **The Ultimate Criterion** 57
Chapter VI. **Student Gain and Instructional Supervision** ... 67
Bibliography 73

INTRODUCTION

The fact that students discuss instructors, that parents question them, and that administrators judge them is not new. Evaluation of teachers has, indeed, a long history, much of it occurring before assessment procedures had become stabilized or even remotely related to theory. The educational literature is replete with discussions of investigations that seek ways of evaluating teacher performance, of predicting effectiveness, and of using various types of ratings in preparing teachers and in improving instruction.

Many investigations of teachers and teaching employ techniques that center about the collection and reporting of demographic data, the teacher's awareness of his discipline, and the teacher as a singular entity functioning independently of his environment. Some studies have been built on an *a priori* approach while others have been conducted on experimental bases that include the application of requisite controls. In most cases, investigations of teachers have been concerned with normative data, personality characteristics, teacher performance, and attempts to relate those variables to success in practice.

A variety of measurement devices, samples, and statistical techniques have been used to study teachers. So-called subjective ratings compete with objective scales for the affection of investigators. Hundreds of investigations conducted over a span of many years in every type of educational institution have failed to suggest a way of looking at teachers and teaching situations that is standardized, replicable, representative of the wishes of the profession, or acceptable to more than one group. A systematic attack on the issue is certainly lacking. However, lack of consensus on approaches to the problem—in fact, a variety of interpretations of the problem itself—has not dissuaded researchers from continuing efforts to appraise teachers.

In many institutions, the appraisal of teachers and teaching usually centers around an activity called, "evaluation of instructors." Although the practice is widespread, it is not universally appreciated. District policies often mandate evaluation of one form or another but just as often, staff members question the techniques employed, the potential use of findings, or the entire process. Acceptance or rejection of the *methods* is often related to the degree of acceptance or rejection of the *purposes* of instructor evaluation. And purposes for conducting such studies vary as much as do techniques for gathering data.

The situation regarding teacher evaluation at any level of education reflects the instability of teacher evaluation in the profession at large. Educational researchers have not been able to isolate dimensions of teachers or the teaching situation which correlate significantly with measures of effect on students or institutions. Hence, it is not surprising that administrators in most schools despair of finding effectual means of evaluating teachers and thus, merely accept or maintain practices that are least likely to stimulate controversy.

Evaluation of college instructors is an important issue with a large and growing literature. Even the university, the foremost bastion of the nonteacher in the realm of education, has become concerned over questions of teaching and teacher effect. Unhappily for the profession, however, it is student disaffection that has triggered the current wave of examination of teaching in the university.

Surveys conducted by the American Council on Education in 1960 and in 1966 studied practices in the evaluation of faculty members in higher education. The first survey obtained responses from 584 institutions, including 25 junior colleges *(49); the second from 1,110, including 128 junior colleges (4). These surveys found much similarity in procedures for evaluating instructors in liberal arts colleges, private universities, state universities, state colleges, teachers colleges, technical-professional colleges, and junior colleges. Similarity in confusion regarding purposes of evaluation and a lack of concern regarding the use of apparently invalid methods to gather data on faculty members were also found. As Gustad noted:

> It was not assumed, when this study was planned, that the situation with respect to faculty evaluation would be found to be good. It was no great surprise, therefore, to find it as it was. What *was* somewhat surprising was the extent and depth of the chaos. . . . It is apparent that little is done to obtain anything that even approaches sound data on the basis of which reasonably good evaluations of classroom teaching can be made. . . . In general, to call what is typically collected or adduced to support evaluative decisions "evidence" is to stretch the meaning of that honored word beyond reason (49).

Astin and Lee's 1966 follow-up of Gustad's survey found little that was different. (4) In evaluating teachers, all or most departments and institutions used "chairman evaluation" and "dean evaluation" —in other words, forms of supervisor rating. Some also collected colleagues' opinions and informal student opinions. The universities and four-year institutions relied heavily on evidence of scholarly research and publication as a measure of teaching. Few institutions of any type reported the use of systematic student ratings, enrollment in elective courses, long-term follow-up of students, or alumni opinions as measures of teacher effect.

---

* Bracketed numbers refer to bibliographical entries on 78-81.

In both surveys, the junior colleges, as a subsample of the group, deviated little from the total findings. However, classroom visits were used by approximately half the junior colleges whereas the total group, including all units of higher education, used these practices to a much lesser degree. Other discrepancies between junior colleges and four-year institutions were that junior colleges relied somewhat more heavily on grade-mark distributions and on follow-ups of students and rarely collected information about scholarly publications. Astin and Lee concluded:

> If the ultimate measure of the teacher's effectiveness is his impact on the student—a view which few educators would dispute—it is unfortunate that those sources of information most likely to yield information about this influence are least likely to be used (4).

Evaluation of instructors is often an inconsistent exercise, archaic, and in large measure, unrelated to apparent purpose. An extensive, recent survey of evaluation practices in California junior colleges revealed nothing to refute that contention (39). Classroom visits; committee consultations in association with deans, colleagues, and division chairman; and ratings provided by students were the chief methods of appraisal. Reasons for the practice were described vaguely as being "to improve instruction," while relationships between procedures and desired results were not made clear.

As generally conducted, research on teachers and practices of faculty evaluation represent two streams of study, both presumably flowing in the direction of something called "better teaching." Curiously, both fail to attend to instruction as a discipline. Rather, they give credence to the instructor in his position as total practitioner—setter of objectives, interactor with students, classroom personality, counselor, scholar, and member of a social environment. And well they must, because the current state of higher education mandates those many roles for its fellows.

Actually, research on teachers and practices of faculty evaluation flow in separate beds. There is a high mound between them and slight evidence of its being cut away. Little has resulted from the millions of dollars spent on teacher-competence research. Practitioners of faculty evaluation often ignore the research and apply their efforts with little understanding of, or enthusiasm for, what they are doing.

The best that can be said for current methods of evaluating faculty in institutions of higher education is that they are ineffectual and little regarded. Not much more can be said for research on teachers at other levels of education. Yet, if research does not pursue questions consistently and eventually produce findings that can be used, it is of little value to the field. And if information gained as a result of research efforts is not seriously considered and incorporated into educational practice, the best one can hope is that teaching does not become worse in the future than it has been in the past.

This monograph is an attempt to reduce the separation between research on teachers and practices of teacher evaluation in the junior college. It examines the current status of faculty ratings, discusses problems in establishing criteria for faculty evaluation, and considers the question of why evaluation should be conducted at all. As its main thrust, it presents a rationale for change. It builds a case for abandoning current practices of faculty evaluation in favor of genuine research on human functioning, on instruction, and on relationships between the two.

The paper is divided into two parts. Part I is a discussion of current practices in faculty evaluation and a report of research in the field. Chapter One reviews faculty rating schemes in current use, examining them from the point of view of instruments and media employed. Problems of rater bias, ambiguous purpose, and above all, indefinite criteria are discussed.

Chapter Two reviews some of the many attempts to relate teacher personality with teaching success. In most of those studies, an *assumed* connection is made between performance and effect—in fact, the words, "performance" and "effect" are often used interchangeably. The assessment of *performance* can be valuable when it builds on theory and furthers knowledge of people functioning in particular work situations. Therefore, although personality appraisal may represent a track separate from assessment of teacher *effect*, it can provide keener understanding of the teacher's role and its potential effect on pupil performance.

In Chapter Three, results of research studies conducted in the U.C.L.A. Junior College Teacher Preparation Program are reported. Variables considered here include personality dimensions of new instructors and ratings of their success on the job. Theoretical considerations and rationale for program design are also reviewed.

Part II presents a case for changing purposes, methods, and criteria of faculty assessment. Among many college educators, the feeling that teaching *should not* be evaluated stems from and leads to a belief that it *cannot* be reliably assessed. The arguments go, "Without clear reason, why do it?" and "Without proper tools, it cannot be done." The two beliefs interact and reinforce each other. Accordingly, Chapter Four discusses both purposes of and criteria for faculty evaluation.

A case for using student gain toward measurable objectives as a major criterion for assessing instructors is presented in Chapter Five. Designs are introduced and results of studies in which teacher effects have been isolated are reviewed in Chapter Six. In addition, changed modes of supervising instruction—hence, instructors—are considered.

Throughout the monograph, the twin issues of "Why study teachers" and "How to study teaching" appear as constant themes. The junior college is seen as an institution which can help in the study of both teachers and teaching by holding to a clear rationale, tying its studies to theory, and participating in genuine research efforts. Current practices of faculty evaluation are seen as being innocuous, at best. They can and must give way to studies that can have a positive effect.

PART I

MEASURING:	Appraising according to a criterion or standard
FACULTY:	People holding academic rank
PERFORMANCE:	A public presentation

THE MEDIA OF FACULTY MEASUREMENT

chapter I

A DEFINITION

All measurement is, in a sense, a means of communication—a medium that allows information to move from one point to another. When thus seen, the many approaches to teacher evaluation become attempts to better understand the performance of individuals as they exercise certain prescribed functions in their various occupational roles.

Media can actually be things (objects) or people (subjects). Teaching machines, audiovisual devices, groups of individuals exploring issues, or a teacher standing in a lecture hall and delivering a monolog are all forms of media. Similarly, psychological instruments of appraisal, pupil rating schemes, classroom observational systems, and supervisor ratings are media. In the first case, the media are employed presumably for purposes of effecting changes in students—of "teaching." In the second case, the media are used to appraise the performance of individuals in special settings. And in both cases, the media communicate information about or from some persons to others, bridging gaps between stimuli and responses.

Although performance measures are frequently viewed as measures of effectiveness, the difference between performance and effectiveness is actually a clear-cut one. The terms do not hold tautological implications. Rather, they suggest two important, separate concepts that may or may not be related. Because the problems of criteria specification very definitely affect both dimensions, it is often difficult to determine the extent of the relationship between them. Many investigators have pointed out that research in the area of teacher effectiveness has been unproductive because of problems associated with the development of suitable criterion variables. It is necessary that these difficulties be recognized so that they can be reconciled. However, concerns about criteria cannot be held as sole reasons for difficulties in teacher assessment. Once specified, criteria must be validated against purposes. The purposes of all educational enterprises center around student learning. Thus effectiveness should be measured only in terms of what eventually happens to the end products—the dependent variables, the students' learning. Faculty performance may or may not be relevant.

TWO MAJOR THESES

This report is based on two major theses: (1) teaching performance as a criterion can be established by such media as supervisor ratings, tests, self and peer evaluations, and observational techniques; and (2) teaching performance must itself be evaluated ultimately in terms of effectiveness. The only valid and stable measure of effectiveness is pupil change—simultaneously, the end product and the single, operationally measurable kind of criterion that can describe teaching effectiveness.

Because such a criterion is significantly absent in most of the research on teachers, it is important to understand what information does exist in the literature and what research findings do suggest. What, indeed, do the many investigations of teaching *performance* (the criterion that is *really* measured) have to say? What are the techniques traditionally employed to assess this dimension? Which procedures should be discarded, or replicated, or directly applied to current operations and implemented in educational institutions" What avenues of study seem fruitful to pursue?

Most of the research on teaching effectiveness or performance has concentrated on the elementary and secondary levels of education. Since research designs can be developed at one level and then extended to others, information about *the* "successful," "effective", "good" kindergarten teacher may be relevant to research on the community college instructor. Some research can be viewed only in isolation. Other studies are pertinent to several situations, their findings equally applicable to populations beyond those immediately considered. These are important considerations to keep in mind while recognizing that, in the extensive material devoted to teaching assessment, the evaluation of college teaching effectiveness is a subject which has not received the critical attention it deserves or needs. Much lip service is paid to the importance of the good teacher, but few criteria for appraising the quality of teaching have ever been established. One reason for the dearth of research and study is that it is difficult to find out very much about what goes on in the college teacher's classroom; traditionally, that place has been sacrosanct and what transpires there exclusively the teacher's business (36).

MEDIA: OBJECT AND SUBJECT

There are various ways in which teaching at either the college or grade-level can be evaluated. For the purpose of this report, media will be divided into two areas: object and subject. Tests, observational techniques, rating schemes, and evaluations forms will be considered object media. Subject media will include the appraisers themselves—colleagues, pupils, supervisors and professors. In this chapter, both object and subject media will be viewed with the major emphasis upon the subject media, the doers who select the appropriate object media for assessing teachers.

The following two chapters will report on techniques and approaches to evaluation of teachers' personality characteristics. This division has been formulated not because personality is a separate entity that can be examined apart from the entire educational system but rather because it is so vast and important that it merits separate treatment. It is suggested that the reader view these reports of teacher performance in terms of the various media and, more important, in terms of their potential value for him in reaching a fuller understanding of the so-called "effective" teacher.

TEACHER EVALUATION: BEGINNINGS

Just as employers in business and industry evaluate their personnel, so teachers are rated by their supervisors and administrators. Formal rating or evaluation practices (in the schools, the difference is actually only one of terminology) stem from efficiency movements of the early 1900's. In 1907, Superintendent Cooley of Chicago noted the tendencies of school officials to give high marks to teachers, and in 1910 E.C. Elliott of Wisconsin released a popular "provisional plan" for measuring teachers' "merits." This plan consisted of a scorecard with several specified areas, each containing subitems that were assigned values and then totaled to arrive at a teacher's score—a form that is still used in many school districts (30).

Since assessment techniques were first reported, a considerable number of surveys have reported the use of teacher ratings. In 98 per cent of the forms surveyed by Boyce, "discipline" was cited as an evaluated quality, with "instructional skill" and "cooperation and loyalty" next in frequency. A rating plan was introduced that included forty-five items listed under five headings: personal, social and professional equipment, school management, and technique of teaching (13).

Later surveys have been described by Monroe (77), Revis and Cooper (94), and the NEA Research Division (81). Many—even those executed forty or more years ago—point to the lack of reliability of existent rating devices. However, the use of such devices still persists. Reavis and Cooper, for example, classified evaluating schemes into several somewhat distinct types, all of which are beset with problems:

> *check scales* which list several attributes, functions or outcomes, each of which is evaluated separately;
> *characterization reports* which report total merit of the teacher according to a scale of values;
> *guided comment reports* which list a series of topics or questions on each of which the rater is required to write a comment relative to the teacher;
> *descriptive reports* which do not specify topics and leave the structure of the statement to the discretion of the rater (as in letters of recommendation);
> *ranking reports* which list the teachers of a given school in order of excellence (94:52).

Many rating methods currently employ these types of reports. The median number of scale values normally used in checklists is five; thus, curiously approximating the "A to F" grading scale!

As early as 1915, the NEA adopted a resolution opposing ratings that "unnecessarily distributed the teacher's peace" (81:63). Recognizing, however, that the teaching profession—as other professions—should evaluate the quality of its services, it insisted that this *not* be done specifically for the purpose of setting salaries. The profession has not yet determined how to "evaluate itself" for the purpose of advancing professionalism without tying those evaluations to a reward system. If the automatic salary schedule is the norm and ratings are viewed as intrusive, the use of evaluation as means of improving instruction will remain minimal.

GENERAL PROBLEMS

In spite of the fact that rating systems are often employed in schools throughout the country—especially with new and nontenured faculty—many questions are raised regarding their usefulness, whether they are practical and, in some instances, even whether they are ethical. On the other hand, value has been claimed. The following has been suggested: It is important for teachers to receive copies of their ratings so that they can improve their performances; tenure recommendations and reappointments of nontenured personnel be dependent upon them; they are useful in selecting individuals for promotion; and pay scales may sometimes be regulated by them. Each of these points may be argued from a "yes, rate" or "no, don't rate" viewpoint.

The real issue in rating schemes, however, is the kind of teacher or teaching held as a model—the criterion against which assessment is made. Scales often fail to define the scope of the teacher's total task in a way that he can orient his efforts and judge his success. Rating forms are often unrelated to the concept of the position or to the objectives of the evaluating schools. Ordinarily it is assumed that the "good teacher" must be a "good person" and thus, any measure that assesses good people *a priori* assesses the good teacher. And although the term "teaching" may be mentioned by investigators as often as "teacher," the whole issue of *effective* teaching is often neglected. Rating schemes generally attempt to rate *people*—often isolated from task, criterion, and the total school situation.

Rating scales themselves have been the subject of study by investigators attempting to determine what characteristics of teachers their administrators consider to be valuable. In the late 1920's Barr and Emans analyzed 209 scales then in use and categorized them by teacher qualities: classroom management, instructional skills, personal fitness, etc. (6). Of interest now is the fact that not only are the major divisions isolated by the investigators the same as those on which rating forms are still built, but also the very wording on many old forms is exactly that used today. Forty years of research on teaching has had little effect even on rating scales used in public school and junior college districts!

In addition to questions about rationale and relevant issues, there are attendant difficulties in measurement. These include the improbability of getting competent observers to evaluate teachers; inadequate samples; distinctions between observation/interpretation and between facts/inferences; relationships of teachers to pupils, colleagues, administrators; and difficulties in training judges.

RATER BIAS

Rater bias is of particular concern in schemes of teacher assessment. There is the possibility of personal bias in any situation where individuals are assessed; when untrained people assess others, the possibility is compounded. People see different people in various lights; one may project his own values and problems upon the assessed without being aware of them. Therefore, any individual who examines evaluations of performance (often based upon unspecified criteria) must also look at the rater to decide from what viewpoint he assesses his subjects. To some extent, this problem may be countered by erecting objective criteria; even so, individuals' biases persist and while they may add flavor to assessment, they may also interfere with it.

Response tendencies—seeing all individuals in a particular light—must also be considered when rating measures are used to appraise individuals. Here, the "halo effect" demands recognition. It is easy to rate individuals in the same way they have been rated in the past and thus to perpetuate failures or successes. The use of superficial assessment techniques (measures that rate teachers' mannerisms, lengths of skirts, or hair), the absence of rater training, and a refusal on the part of many instructors to accept ratings as a necessary concomitant of professionalism are other issues demanding recognition. Indicative of the magnitude of the problem was Barr's statement that the safest approach to the appraisal of teaching is a multiple one "employing more than one theoretical orientation, a variety of data gathering devices, and . . . a number of persons studying teachers and teaching under a variety of conditions" (5:28).

After well over a half century of efforts, we are still at the most rudimentary empirical stage of assessing instructors. Because many studies have indicated low correlations among such variables as supervisory ratings, pupils' gains, and instructor-examination results, the field might better concentrate on the *products* of learning and teaching rather than on isolated, sometimes irrelevant dimensions.

A more recent study of teachers has attempted to bring research tools to bear on assessment. In his vast studies of teacher characteristics, Ryans' (107) was concerned with issues of unreliability, bias, and indeterminate criteria. He developed instruments to record classroom activities of teachers; sought to determine major patterns of observable teacher behavior; and developed questionnaires to tap such characteristics as attitudes, educational viewpoints, verbal intelligence, and emotional adjustment. Focusing particularly on the stabilization of teachers' classroom activities, Ryans found that observer training was an essential preliminary step to correlating behaviors with teacher characteristics:

> Only with training of observers can one expect to obtain meaningful assessments of teacher behavior. It is the only proper way one can approach teacher assessment for either research purposes or for preservice and inservice teacher evaluation (107:74).

Reliability was increased by such procedures as focusing observers' attention on a limited number of behavioral dimensions and providing

> . . . specific and unequivocal operational definitions of the characteristics to be assessed; intensive training and practice sessions with observers; immediate assessment of each specified behavior; care to avoid rating biases; replication of observations by independent similarly trained observers (107:75).

However, even after training the observers, interrater reliability coefficients, while high, did not approximate the optimal.

Measures of evaluation that are usually more objective than classroom observation systems also fall short of desirability. It cannot be new to anyone who has gone beyond elementary school that different teachers—using the powerful tools that are grades—rate students on the bases of varying measures. Johnny may receive an "A" for a paper that is ostensibly like the one for which Mary was given a "C." Miss Jones and Mr. Brown may consistently give polarized marks to the same students who submit similar work. This same kind of discrepancy also appears in the grading of student-teachers by supervisors and training specialists. And here, too, the concepts of projection, individual differences, personal biases—even whether the rater slept well the night before—enter into the picture.

Rating forms filled out by supervisors, colleagues, pupils, or administrators provide a somewhat more objective approach to assessment. These, too, are subject to many of the problems already designated and are compounded by ambiguity in definition of concepts and criteria.

Other types of assessment media are similarly fraught with difficulties. The most valid and reliable published tests (of personality, achievement, attitudes, for example) are subject to misinterpretation, problems of which the would-be rater of potential investigator must be aware. *The Handbook of Research on Teaching* provides a good basis for appraising both subject and object media (41).

SUPERVISOR RATINGS

Perhaps the oldest and most practical approach to teacher evaluation is through supervisor ratings. Indeed, it has been suggested that in spite of the many predictive efforts based on ratings made by students, colleagues, supervisors, or independent researchers, evaluations by campus supervisors consistently prove to be the best available yardsticks for predicting success of neophyte teachers (83).

Some problems attendant to this type of evaluation were cited earlier in this chapter. In a later chapter, studies of junior college teaching interns will be discussed, and correlations between their psychological instrument ratings and their supervisor evaluations will be presented. Since these investigations deal rather thoroughly with the use of supervisor ratings, such media will not be discussed at this time.

RATINGS BY DEGREES

Other attempts to evaluate teachers' performances have been concerned with such objective variables as types of attained degrees and size and kinds of degree-granting institutions. The literature devoted to teacher assessment and the prediction of teacher success is replete with discussions of such measures. In the section on the junior college in the *Encyclopedia of Educational Research* (53), much attention is given to the question of the academic degree. This is consistent with many recent surveys that report such information on junior college faculty members but evoke little, if any, reference to data that are nonnormative in nature.

As junior colleges move closer toward the goal of minimum standards that include a master's degree for most faculty members, dimensions other than those usually classified as demographic or normative (sex, age, education) need be considered. What characteristics significantly discriminate between effective and noneffective teachers? The degree itself seems to be only one part of the question and will not yield much information toward understanding individuals or appraising faculty. What are the related dimensions? Which correlate most significantly with teacher effectiveness? Which predict the most "successful" teacher?

RATING BY COLLEAGUES

Other "subject media" in teaching-rating schemes are faculty members who appraise the performance of their associates. Such procedures are often informal and undocumented as to value although, on the surface, they seem to contain some merit. A suggestion that one may invite "colleagues whose judgment he respects to sit in on his classes and to provide critiques of procedures and relationship observed" (36:2) also includes difficulties. The feedback that peers offer is conceivably valuable but, like the "round robin" exchanges that typify certain adolescent searches for self-knowledge, it is likely to be fraught with subjective, nondirective assessments. Sharing ideas may have short-range value but in extreme cases, it can prove harmful to the teacher and his teaching situation. The profession will not develop out of "tips" and unvalidated techniques. However, evaluation by colleagues has one advantage—it is the scheme least likely to meet with resistance.

SELF-EVALUATION

Perhaps the most difficult—and eventually, the most rewarding—kind of evaluation is evaluation of self. This assumes both a degree of maturity and a need for objectivity—difficult for all to attain and for some, impossible. As a way of looking into one's self, introspection may result in self-evaluation. On the other hand, self-examination may become a circular route to nowhere if it implies only such questions as "How am I handling the students? Did they like the lecture? Am I in a rut? Do I stimulate their creativity?" Even with self-examination, there needs to be a definite basis upon which one must structure goals and objectives, and which acts as a criterion for careful scrutiny.

Brown and Thornton (17) take the position that college teachers can evaluate themselves by such procedures as: introspection; studying the product; asking colleagues and student committee members to sit in on classes and evaluate; recording class sessions; noting the extent of student participation, the quality of their comments, and the types and qualities of the teachers' own comments; and finally, collecting student ratings. They further suggest that the instructor should not average rating forms submitted by students but rather, should note patterns of responses that cluster about particular strengths and weaknesses.

The American Association of Colleges for Teacher Education furthered efforts at self-evaluation by publishing a list of seventeen "teacher self-evaluation tools" (113). These items were listed in order by 5,303 college and university instructors who had found them valuable. It was found that such procedures as planned meetings with colleagues and the taping of regular class sessions were particularly helpful. More important than the items in this list of seventeen was a statement of the uses to which they were to be put. Where ratings were to be made by colleagues, students, administrators, or by self, the instructor alone determined what he would do with the findings. Unfortunately, however, no attempts were made to follow up the effect of this list by determining whether anyone had changed his teaching practices as a result of what he learned about himself.

EVALUATION BY CONTRACTS AND GRANTS

A unique approach to teacher evaluation is through the determination of publications and government awards attributable to faculty members. A preliminary report of an investigation at Tufts University suggested that, contrary to the opinion of many who are engaged in teaching/research struggles, publications and awards *do* relate to ability in teaching undergraduate students.

Although many statements in the popular literature and in professional journals suggest that publication efforts and government support for research activities detract from teaching effectiveness in the classroom, the Tufts data do not support these conclusions. Rather, it was found that students rated as their best instructors those faculty members who had published articles and received government grants and/or other support. Whether this is true of the faculty in other higher education institutions is a topic certainly worthy of further investigation. For example, faculty members in community colleges are becoming more frequently involved in extramural-funded projects. Is there a relationship between leadership in such endeavors and "good teaching"?

STUDENT EVALUATIONS

Evaluation of instructors by their students has been a popular practice for a number of years. In fact, in spite of a somewhat cynical opinion among some teachers that very little value can be placed on student judgment, greater attention is now being given to student ratings than ever before. Questionnaires, checklists, and rating forms have been used extensively by students at different levels of education and in hundreds of school settings. Stecklein (116), for example, reported that of 800 colleges, student ratings were regularly used in nearly 40 per cent and that an additional 32 per cent were considering their use at the time he conducted his survey.

Ratings by students, of course, are subject to many of the same criticisms that relate to other measures of judgment based on nebulous criteria. Some investigators report they are stable and reliable means of assessment (50), while others point out that the level and size of class significantly relate to students' opinions. The "role concept" held by the student—that is, the "image" he holds of what a teacher ideally should be—also undoubtedly influences the ratings.

McKeachie has many times surveyed the field of college teaching, has conducted studies of his own, and has reviewed and analyzed the work of others. His conclusions are worthy of note in the context of student ratings, as are the gaps he describes that must be filled before teaching effectiveness can be assessed on any scale worthy of mention. One gap is that very little is presently known about what college teachers do. The usual sources of information are students' complaints and colleagues' impressions, not always substantive or valid.

When student ratings are utilized as a way of bringing order into the communication process among students, faculty, and administration, a second gap in knowledge becomes apparent. We do not yet know how well students can rate teaching effectiveness. Although the few studies that have assessed that dimension seem to point to the fact that they know when they are being well-taught (78), there is not enough available evidence to be certain of what "well-taught" implies.

A third gap is that "even if we can measure some aspects of teacher behavior validly, we do not know the relationship of that behavior to student learning, which is one of our ultimate criteria of effectiveness" (71). This, of course, is the crux of the matter—the relationship of teacher behavior to student learning is not known and, despite decades of research, we have not yet begun to understand those influences.

McKeachie's own work includes the observation of instructors by students who rate the degree to which they performed in classes. Students' and trained observers' ratings showed high correlation on certain dimensions, not on others; however, McKeachie agrees with the assumption that students can rate instructors accurately (76).

Ratings given by students to courses tend to be more highly correlated with their own achievement than ratings given to instructors although it is often difficult to separate the two in the students' minds. In one study, data were collected on 87 instructors from 4,285 student scales. The major questions investigated asked:

1. Does students' sex, age, major, level of education, grade-point average, or course grades previously received from the instructor being rated have any relationship to ratings of instructors?

2. Are instructors who differ in sex, age, faculty rank, degree held, major area, or length of teaching experience rated differently by students? (93:4).

Students rated their instructors on each of three seven-point continuums based on the traits of behavioral patterns identified by Ryans (104):

I. aloof, egocentric, restricted behavior vs. friendly, understanding behavior
II. evading, unplanned, slipshod behavior vs. responsible, systematic, business-like behavior
III. dull, routine behavior vs. stimulating, imaginative, enthusiastic behavior.

A short instructor characteristics scale was also administered in order to obtain information from faculty regarding their age, sex, rank, degrees, major area of concentration, and length of teaching experience.

Analysis of the data suggested that:

1. Students remembered and accurately reported their grade-point averages. The correlation for 100 randomly selected student-reported GPA's (grade-point average) and their actual GPA's was .96.

2. Student ratings of instructors were not substantially related to student's sex, age, grade-level, major area, grade-point average, or grade(s) previously received from the instructor they rated.

Ratings were different among various departments within the school of education and, further, younger instructors with lower faculty rank and fewer years of teaching experience were consistently rated as more "flexible, understanding, systematic, responsible, business-like, stimulating, imaginative, and enthusiastic" (104).

Students at St. John's River Junior College evaluated their instructors in 500 classes on the basis of four measures:

1. Positive personal traits (appearance, attitude)
2. Scholarship
3. Skill of presentation (knowledge)
4. Accuracy in evaluating students (understanding, fairness) (84).

A rating form was devised whereby students could rate instructors on the basis of a five-point scale for each quality. The study was conducted by the dean of the college who personally visited each classroom and passed out copies of the rating form to each student. Then both dean and instructor waited outside the door for the students to complete the ratings anonymously. Most of the faculty was in favor of that procedure.

Knapp and Goodrich (66) studied college teachers by interviewing them and by collecting ratings from their former students. Factor analysis of qualities that correlated with high effectiveness in motivating students pointed to three general characteristics: masterfulness (shown by severe grading standards, leadership in departmental enterpreneurship); humor; and intellectual distinction.

Guthrie (51) found that length of teaching career did not correlate with the merit ranking of teachers and that it was not a significant factor in ratings. His extensive survey of students and faculty ratings of teaching effectiveness suggested that the faculty tended to rate scholarly attainment as important, while students accented personal qualities in teaching. Students were not found to give higher ratings to teachers within their own fields of specialization.

In terms of the three designated functions of college teachers—research, informational and character-developing—Knapp concluded:

> . . . professors tend to esteem and respect themselves primarily on the basis of their research function. Students and their administrators, however, especially in smaller institutions tend to value most the informational and character-developing functions (and) . . . the public at large is probably inclined to attach great significance to what we have called the character-building function. Thus different segments of the population, to whom the college professor must in some degree answer, apparently expect different kinds of performances (65:306).

Problems in the collection of student ratings result from many variables; for example, their use is affected by course size. In an analysis of student ratings at the University of Illinois, teachers in courses with thirty to thirty-nine students consistently received lower ratings than did those in courses with either more or fewer students. Teachers of on-campus courses received worse ratings than did those of off-campus courses; teachers of electives were rated more favorably than were teachers of required courses.

The argument that a single standard alone is dangerous to use may be raised here. Why are instructors in very small and very large classes rated higher? Are such sections easier to teach? Or are the students more lenient in their ratings? Similarly, are teachers more effective in elective courses or are the students easier to please because they are better motivated? If only student ratings are available, such questions are unanswerable.

One way around the difficulty would be to have separate tables of norms for interpreting student ratings in all course categories and then to compare the ratings earned by one instructor against the norm. If for no other purpose, student ratings may be used for teachers' self-improvement in the sense that instructors can establish their own norms over the years. Student biases may still be operative but if instructors interpret the ratings for themselves, they can sometimes obtain information valuable for diagnosing their problem areas by noting clusters or profile formations.

In a study by Cooper and Lewis (27), the Rorschach was used to assess teachers who were independently rated by students on a checklist. Teachers who were considered to have good student relationship tended to possess such personality traits as a sense of humor, courtesy, tact, fairness, flexibility, self-control, ability to create interest, sympathy, friendliness—and on and on. Such general qualities of "goodness" only seem to emphasize again that "good teachers" are "good people." This type of research offers little more to what is already known.

A series of teacher self-appraisal instruments to which students may respond has been developed by Simpson and Seidman (113). Evaluation items are designed for several areas, including open-ended questions, checklists, and rating scales, complete with 291 illustrations from which the teacher may select. As an example, included among the open-ended illustrations for Area I, General Course Evaluation, were:

- What were your most stimulating and challenging experiences in this course?
- This course would have been more valuable to me if:
- What are the *one* or *more* least satisfactory features of the course?
- How would you rate this course in comparison to your other courses? Why?
- What did the teacher fail to do that you felt would have been beneficial to you?

The source is particularly valuable for instructors desirous of constructing their own rating instruments.

WHAT IS BEING RATED?

Object media—tests, rating forms, observational techniques, and similar measures for evaluating teaching performance—have been cited in this chapter. Discussed more thoroughly have been subject media—the actual doers or appraisers who, by employing certain techniques and juggling certain variables, attempt to evaluate teachers. Student ratings have been reviewed in particular depth because they can be shown to relate, at least in principle, to the ostensible purposes of educational institutions.

In many cases the subject media—supervisors, self, peers—have employed an *a priori* approach to assessment. In others, they have carefully scrutinized the available material and proceeded to appraise individuals on the basis of clearly defined schemes. In most cases, however, conceptual frameworks have been ambiguous—if they have been defined at all; and even the more consistent studies often ignore the fact that while teachers are people, all "good" people may not be "good" teachers.

Some patterns in teacher assessment are apparent, particularly the fact that in almost all measures of faculty—whether subject or object media are employed—it is the teacher's *performance* that is being assessed. The distinction between *performance* and *effect* must be borne in mind when any study of faculty is being discussed or reviewed. The connections, if any, between the two are as yet thoroughly unclear. But performance is measured, examined, and evaluated as assiduously as though it were the end goal of every public educational structure! Although "performance" is a severely limited definition of the term "teaching," it is usually accepted in the schools as being a sufficient condition for teaching.

Another persistent problem in the field is that theoretical constructs are often confused with observational descriptions. Postulates assumed to underly behavior are mentioned as though they were the behavior itself. For example, "teacher competence," a quality dependent upon interpretation, cannot be observed directly. It can be inferred from descriptions of teacher actions, yet the term is often used as though the construct could itself be observed.

A more meaningful definition of "teaching" and a tendency to speak of constructs in operational terms are necessary first steps to critiquing schemes of faculty evaluation. Until the desired traits of a teacher are decided upon, no comprehensive definition of teacher worth is possible. Research on teacher characteristics and evaluation of faculty members can have definite impact upon schools only if there is agreement upon terms used and upon definition of variables for which the terms stand.

The problems in assessment are legion. Some of them have been presented here and others will be cited in the next chapter in conjunction with a survey of certain teacher-effectiveness studies that attempt to view teachers in terms of personality variables.

EVALUATION THROUGH PERSONALITY VARIABLES

chapter II

Evaluation of human performance almost invariably incorporates, either by direction or by implication, evaluation of personality dimensions. Thus, a preponderant number of investigations dealing with effective or noneffective performance are actually concerned with subjects as individuals—their special characteristics and traits—although personality appraisals may not have been considered as essential parameters of the original research designs. Most of these studies equate certain personality characteristics with teaching effectiveness.

BACKGROUND

Approaches to individual understanding through personality assessment are neither unique nor recent phenomena. Rather, they have a long history, stemming from the early Greek scholastics who attempted to measure people by categorizing them as specialized types or "humors" and continuing to our current, relatively mechanized modes of perceiving man (14). People have been examined as single individuals, as deviates from established norms, and as members of various types or subgroups.

Until Murray (79) proposed his individual need/environmental press concept, little attention was paid to the interactional effects of people as they function and relate to others in particular situations. The concepts of projection and observer roles had appeared earlier but the integration of situational and personal determinants in predicting performance had not been previously emphasized. These variables have since been measured by the extensive work of the Office of Strategic Services assessment teams, the many college environmental studies conducted by Pace (85; 86) and Stern, Stein and Bloom's recognition of the need for consistency between:

> The frames of reference of the original assessors and those individuals in a field situation who would be asked to appraise the subsequent performance of the assessees (117:28).

Partially as a result of these activities, individuals today are perceived both as unique entities and as they interact with significant others in prescribed settings.

Also within the last few decades, personality assessment has turned from a predominantly clinical orientation that stressed pathology and deviation to appraisal of individuals in groups. Eiduson (35), Hahn and MacLean (52), Roe (97; 98), and Super and Crites (118) are among the many who seek to understand people as members of special occupational forces.

The field of education and the people working in it, especially the faculty, have accounted for a considerable amount of research. In this area, generated by the many questions that pertain to individual performance, evaluations of teaching performances often have been interwoven with assessment of teachers' personalities. To this end, multifarious measures have been employed, measures which stem from one or more theories of personality, or in many cases, from none at all (42). Independent ratings of both experienced and beginning instructors range from simple value judgments to elaborate questionnaires and intricate statistical procedures which isolate a range of pertinent parameters. Variables extend from the singular to the most sophisticated, from the simplest to the most complex. They embrace populations that vary from small and homogenous to large and heterogeneous.

The research has attempted to answer questions in many ways. For example, specific personality characteristics have been isolated and then plotted against estimates of future teaching success. Such variables as belongingness, empathic potential (31), and organization (21) have been posited as traits to discriminate between successful and unsuccessful individuals who adjust to new situational demands with varying degrees of success. Both general and global judgment of personality [16; 128] have also been used to predict success, Cohen and Brawer suggesting accordingly that:

> Judgments of global personality may well provide preferable means for assessing both the general adjustability of teachers and teachers-in-training and their ability to integrate past experiences with present situational demands (16:180).

Dimensions of flexibility and rigidity have been employed to determine the openness or closedness of belief systems (45; 99), while a variable described as "cognitive flexibility" has been examined so that operational translations of intern-teachers' behaviors might be derived.

The major criticisms of all this research, repeated over and over, deal with the lack of independent criteria upon which to base appraisal and the "theoretical vacuum" in which so many studies are conducted. Getzels and Jackson, in particular, point out that:

> Despite the critical importance of the problem and a half-century of prodigious research effort, very little is known for certain about the nature and measurement of teacher personality, or about the relation between teacher personality and teaching effectiveness. The regrettable fact is that many of the studies so far have not produced significant results. Many others have produced only pedestrian findings. For example, it is said after the usual inventory tabulation that good teachers are friendly, cheerful, sympathetic, and morally virtuous rather than cruel, depressed, unsympathetic, and morally depraved. For what conceivable human interaction—and teaching implies first and foremost a human interaction—is not the better if the people involved are friendly, cheerful, sympathetic, and virtuous rather than the opposite? What is needed is not research leading to the reiteration of the self-evident but to the discovery of specific and distinctive features of teacher personality and of the effective teacher (42:57).

However, all studies are not subject to these criticisms. Many are built upon defined criteria and defined concepts, consider the need/press rationale, and could be reasonably well replicated for different levels of education. Others have contributed important information about teachers and teaching behavior and thereby aid in the general evaluation of education. And still other studies may be potentially useful if they can be viewed in terms of global concepts or specific measures that are related to criterion variables.

Because many investigations of teacher personalities have been extensively surveyed in the *Handbook of Research on Teaching* (41), *The Encyclopedia of Educational Research* (53), and elsewhere, this monograph will not attempt to present a comprehensive review of the literature; instead, it will cite a few selected studies. Some recent research on junior college teaching interns that deals with personality appraisal and its relationship to supervisors' evaluations of teaching success will also be presented in following chapters.

A SAMPLING OF REPRESENTATIVE STUDIES

Generally those studies dealing with characteristics of successful teachers have been concerned with the collection of opinions from experts in teacher education, students, teachers themselves, laymen, and administrators. Little has been done to determine the importance of these characteristics by definitive techniques. Ryans' (106) extensive research in this domain did isolate some dimensions relating to the effectiveness of teaching performance. However, while his observational reports described major patterns of classroom behavior, his results were generally disappointing. Reporting results of the National Teacher Examinations for teachers of different grades and various subject matters, Ryans presented consistently dissimilar profiles (106)—outcomes that make us wonder whether it is at all valid to measure elementary school teachers on the same bases that we would assess college chemistry instructors. Of further concern is the fact that while many of the appraised characteristics were very general (for example, that the teacher must be "understanding," "sensitive," "have empathy"), Ryans found that teachers who were rated as "egocentric" seemed to perform as effectively in their work as those judged otherwise (105).

Dugan also studied the relative importance of selected factors in the effective teacher. She designed a questionnaire to measure such dimensions as egocentricity, mental objectivity, extraversion, and introversion. The instrument consisted of such questions as whether a new teacher joining the school faculty is "taken under your wing" and if the teacher discusses questions regarding grades with his students. The results of this research suggested:

> that unselfishness or emotional stability has not been proved to be necessary for effective teaching comes as a surprise only to educators. Pupils seem to have always realized this. In fact, pupils are surprised that some teachers are every bit as emotionally mature as other people . . . or in other words . . . normal (33:337).

As Getzels and Jackson (42), so strongly point out, the specification of a criterion of effectiveness has been a major stumbling block in research on teaching. Traditionally, there have been three ways of establishing the criterion:

> First, from an evaluation of the scholastic achievement of students; second, from ratings or judgments by supervisors who have observed the teacher in the classroom; and third, from ratings or judgments furnished by the students themselves (47:119).

PERFORMANCE

If performance is considered to be an acceptable criterion of "teaching"—and it often is—there appear to be techniques and instruments available "which can provide acceptable indices" of behavior (47:120). For example, Durflinger (34) built a 41-item teacher-evaluation scale and Michaelis (75) developed a form that yielded quantitative indices of teacher progress. Veldman and Peck's (126) 38-item pupil observation survey yielded five factors measuring class room behavior: friendly, cheerful; knowledgable, poised; interesting, preferred; strict control; and democratic procedure. Factors I and IV (friendly, cheerful, and strict control) appeared to correlate highly with supervisors' ratings of teachers while Factor II had a moderate correlation. It must be emphasized, however, that these and other studies which will be cited in this connection are based on the premise that performance is tantamount to effectiveness.

The application of psychological tests as predictors of teaching effectiveness has been carefully reported in Gage's *Handbook* (41) and in many other surveys of research. These measures include both objective and projective inventories, new and established, that report varying degrees of validity and reliability. The most widely used instrument for the measurement of teacher attitudes has been the Minnesota Teacher Attitude Inventory (MTAI), developed by Cook, Leeds, and Callis (26) from their research on teacher attitudes toward children. It is designed to measure those attitudes of a teacher which will predict how well he will get along with pupils, his interpersonal relationships, and indirectly, how satisfied he will be with teaching as a vocation. Norms are presented for high school students, college students, teacher trainees, and experienced secondary and elementary school teachers. For preliminary tryouts on the MTAI, (69) criterion groups were established by asking principals of seventy elementary and secondary schools to designate several teachers whom they considered to be superior and several considered inferior. Identifications were made on the bases of (1) the teacher's ability to win pupil affection; (2) his fondness for children and understanding of them; and (3) his ability to maintain a desirable form of discipline.

In spite of its popularity, research with the MTAI has not always lived up to expectations (108; 19). However, new studies and replications of existing research may well establish more encouraging results as, for example, those reported by Tanner (121) who noted considerable differences in expressed values between the sexes. He found that at least two of the value-areas (economic and social) seemed to describe groups of teachers who were judged to be "superior" and "inferior." And Seagoe (111), correlating student teachers' ratings with their ratings by principals of field success three years later, found that economic and aesthetic values were most highly related to ratings of effectiveness. However, neither value was consistently related to the designated criterion of success.

The Minnesota Multiphasic Personality Inventory (MMPI) (55), a clinical tool widely applied to nonclinical situations, has also been used in teacher appraisal. Here, too, the research evidence is often discouraging (125; 75), although moderately positive results have been reported with the use of a "sign" approach (46).

Cattell's 16 Personality Factor Test (16PF) (19) was employed in a study that attempted to answer questions regarding the presumed relationships between personality and teaching success. Lamke (67) tested 146 University of Wisconsin students enrolled in undergraduate education courses in the psychology of learning. His criteria for teaching success were the opinions of "experts" about the subject's student teaching performances and their acceptability by the principal or school superintendent. The data suggested that good teachers are more likely than poor teachers to be "gregarious, adventurous, and frivolous"; to show emotional responses and strong artistic or sentimental interests; to be interested in the opposite sex; and to be polished and "cool." Poor teachers were more likely to be "shy, cautious, and conscientious"; to lack emotional response and artistic or sentimental interests; to have comparatively slight interest in the opposite sex; to be easily pleased; and to be more attentive to people.

Other techniques used to predict teaching effectiveness include the Heston Personal Adjustment Inventory (56), the Minnesota Personality Scale (29), the Minnesota T-S-E Inventory (37), and the Rorschach (120). These have shown varying degrees of predictive and concurrent validity; for example, Cohen and Brawer found that Rorschach assessments for global personality adjustment and integrative ability:

> . . . show a high correlation (.57 and .61) with college supervisors' ratings. They correlate less well (.44 and .54), but still significantly, with the ratings given to (junior college teaching interns) . . . by the campus training director (24:184).

Travers (123) found no relationship between adjustment scores derived from the Monroe (77) Rorschach checklist and desirable teaching behavior (as indicated by administrators) although a specific pattern of Rorschach scores did discriminate between highly desirable and undesirable student teachers. For the more highly rated student teachers, this pattern reflected a strong need for achievement and emotional outgoingness or extraversion as manifested by their orientation to environmental stimuli.

Another psychological instrument used to predict teaching performance is the California Psychological Inventory (44), developed

> . . . to assess "folk dimensions" of interpersonal and interactional behavior . . . such as socialization, psychological-mindedness, and flexibility (47:120).

Durflinger (34) and Hill (57; 58) both conducted studies with the CPI. Gough, Durflinger, and Hill (47) used a sample of female students from the University of California, Santa Barbara, who were doing supervised classroom teaching. Combined supervisor ratings were used as the criterion of performance. Since correlations of these ratings with the CPI scales for flexibility (Fx), good impression (Gi), sociability (Sy), and socialization (So) were modest, a combination of scales to provide a more useful basis of prediction than any of the single scales was sought. Analysis of the eighteen scales again the teaching criterion gave rise to an equation which included Sy, So, and Py (Psychological-mindedness) as positive weights.

Further analysis of these scales was conducted with 124 females and 78 males in an instructional program at Ball State University who were dichotomously grouped as superior or inferior in teaching effectiveness. Analysis of the equation derived from the CPI scales suggested:

> . . . personological bases of conscientiousness vs. under control of impulses for males, and of resoluteness versus indifference for females (47:119).

Male students scoring high on the CPI equation were described on the Adjective Check List (ACL) (45) as follows, with the words listed in order of magnitude of correlation with an independent measure: conscientious; practical; rational; moderate; methodical; planful; responsible; logical; reasonable; capable; thorough; reserved.

The low-scoring male was described as: reckless; daring; pleasure-seeking; spendthrift; irresponsible; flirtatious; show-off; spontaneous; adventurous; mischievous; quick careless.

The words used most differentially to describe high-scoring females were: dominant; persevering; persistent; serious; opinionated; ambitious; demanding; logical; rigid; clear-thinking; determined; responsible.

And finally, low-scoring college women, who also scored low on the CPI equation, were described on the ACL as: curious; affectionate; careless; easy going; unconventional; dreamy; understanding; irresponsible; cheerful; natural; individualistic; thoughtful.

What, then, do these various sets of twelve descriptions say about the four groups of scorers? The high male scorer may be described as a

> . . . diligent, effective, individual, well-organized, attentive to the practical demands of his work, and thorough and conscientious in carrying out his duties . . . self-disciplined and reserved, not at all flamboyant or unconventional . . . the kind of person who can be counted on to display discretion and good judgment in any situation (47:124).

On the other hand, the low-scoring male's description suggests a clearly evident "syndrome" of behavior and temperament. On the CPI equation for forecasting teaching effectiveness, he appears to be

> . . . undercontrolled, unbridled, too much dominated by his own impulses. Although in many ways an attractive personality (spontaneous, adventurous, quick), and probably original in his perceptions and ideas, he is too irresponsible and too careless to perform effectively in a day-by-day classroom situation (47:124).

The women identified by the equation as potentially effective student teachers are seen as being quite different from the men so identified:

> . . . the high-scoring young lady is a strong and resourceful individual, clear and explicit about her goals, and resolute in pursuing them . . . her seriousness of purpose and determination are such that those who know her well find her somewhat rigid and opinionated, however worthy her ambitions and steadfastness (47:125).

The low-scoring college woman, on the other hand, is described somewhat like the low-scoring male, although the flavor of the cluster differs. She is

> . . . somewhat undercontrolled . . . but . . . affectionate, thoughtful, and of an optimistic turn of mind. Hostility, aggression, rebelliousness—all qualities which one might hypothesize as negatively related to teaching effectiveness—are alien to the pattern actually delineated. Our low-scoring S seem very likable, easy to get along with, a pleasant and undemanding individual. But as a teacher she will not do; her lack of organization, overresponsiveness to distractions of the moment, and indifference to practical realities are drawbacks too great to be ignored (47:125-26).

The criterion problem, central in the minds of many contemporary researchers, was seen in a relatively different light by Dandes (28) who suggested that, since educational goals are seldom considered in the selection of criterion for teacher effectiveness, much of the research is accordingly inconclusive. If, on the other hand, variables are examined in the light of specified educational goals, there may be certain consistencies. And if these goals include such objectives as growth and self-directedness, personal responsibility, spontaneity, critical problem-solving ability, etc., then certain teacher characteristics may be associated with student development in these particular directions.

Assuming that dimensions of psychologically healthy (rather than psychologically pathological) attitudes and values of teachers may be related to their role effectiveness, Dandes (28) used several instruments to tap these dimensions: The Personal Orientation Inventory (89) was used to measure general health or psychological well-being; the Minnesota Teacher Attitude Inventory (26) rated dimensions of permissiveness and warmth; authoritarianism was measured by the California F scale (1); the openness-closedness of belief systems was defined by scores on Rokeach's (99) dogmatism scale; and the LC Scale (70) rated the liberalism and conservatism of educational viewpoints held by the 223 subjects. Multiple correlations suggested a significant relationship between measured psychological health and the specified attitudes and values of the teachers. The greater the psychological health—as defined by the author—the greater the possession of attitudes and values deemed characteristic of effective teaching. Scales of liberalism and permissiveness were positively related to psychological health; authoritarianism and dogmatism were negatively related. This investigation reported that subject information or knowledge of teaching techniques alone does not insure teaching effectiveness and that, in fact, a teacher may possess all possible knowledge and still be unable to communicate in a "psychologically healthy" framework.

Also examining the relationships of teaching effectiveness and the teacher's personality, Symonds tested the hypothesis that the

> manner of teaching is an expression of the teacher's basic personality reactions, and that these reactions constitute the core of teaching behavior in the classroom situation (119:180).

A small group of subjects—fourteen females and five males—whose teaching experiences ranged from less than five to more than thirty-five years in elementary, junior and senior high schools, responded to three kinds of evidence: tests, interviews, and observations. The Rorschach technique and the Thematic Apperception Test (TAT) (79) were administered. An average of ten interviews focusing on attitudes were conducted and classroom interactions were observed. While the relationships reported cannot be considered definite (in terms of the limited population), the conclusions do demand consideration by anyone interested in training, selecting, and evaluating teachers at all levels of education.

Among the difficulties described as rendering teachers ineffective, Symonds (119) cited feelings of inadequacy, insecurity, or inferiority. Such responses as overaggression, bluntness, bossiness, unfeelingness, and snappishness may constitute reactions to these feelings. And since such feelings are neither learned nor unlearned in teaching courses but are, rather, the outgrowths of basic patterns, they should be subjected to procedures that encourage change. For example, support and encouragement provided by colleagues and administrators during an individual's first months as a teacher may provide opportunity for change. Another is to become sufficiently familiar with various phases of the teaching roles so that "normal" feelings of inadequacy may be met by positive help situations. If feelings of insecurity persist, other measures such as counseling, therapy, or group training may be needed. In any case, since expressions of hostility, growing out of aggressive feelings and projected onto students, appear to be basic personality dimensions that undermine the teacher's performance, they must be confronted.

Conversely, docile, easy-going attitudes may also be reactions (actually, reaction formations) to feelings of inadequacy that result in ineffective behavior. They similarly demand attention. Dependency needs may result in outward behavior wherein the teacher assumes many responsibilities beyond his abilities. In some teachers, slow or dull pupils may evoke threats of failure associated with painful childhood memories.

Teaching is at least in part an expression of personality; accordingly, it is important that a teacher be free to adopt procedures consistent with his basic attitudes and perceptions. Although methods and procedures learned during the teacher's college preparation may superficially influence his behavior, they do not actually determine the nature of his relationships with pupils or peers.

Peck (87) attempted to predict ratings by school principals from the independent analysis of personality data on instructors in five different Texas school systems.

All forty-nine subjects were experienced elementary teachers who were rated by their principals on the basis of the Teacher Effectiveness Scale, a five-part criterion instrument. The school principals were then asked to nominate teachers who were high, average, and low on five scales:

I. Organizing and communicating information and skills
II. Creating a healthy realationship with pupils
III. Creating good relations with other teachers
IV. Building good relations in the community
V. Supervisor's personal evaluation ("Who would you pick to take with you if you moved to a new school?") (87:70).

All nominated teachers then completed two forms: a biographical information form and a ninety-item sentence completion test developed by Peck. Three highly significant correlations suggested that there may be some "fairly universal standards" for judging the effectiveness of teachers, at least at the elementary level; that various principals offer stable and valid judgments; and that personality may, indeed, be appraised from projective data.

Guba and Getzels (47) found reliability among raters in a study of teacher effectiveness conducted among military personnel. Although the purpose of this investigation was not to discover relationships between personality and effectiveness, the data supported such relationships. Using Fisher T tests, Rosenzweig (102) found that: (1) extrapunitiveness (the tendency to blame the environment) was linked to teaching ineffectiveness; (2) intrapunitiveness (the extent to which aggression is turned by the subject upon himself) was linked to effectiveness; (3) inpunitiveness (extent to which aggression is evaded) was not significantly linked to effectiveness; (4) obstacle dominance (responses in which barriers that overcome frustration are predominant) is related to effectiveness for some groups of people and not for others, as are (5) ego-defensiveness and (6) need persistence (emphasizing solutions to problems).

The degree of abstractness or conceptual level (CL) that a subject employs (110; 62) has also been examined. Individuals who demonstrate higher CL levels are expected to be more flexible, more able to tolerate stress, and more capable of offering alternate solutions than individuals with lower CL's. These dimensions were all considered important in the effective teacher but, like many other variables, they do not necessarily describe unique teachers but rather so-called, well-adjusted "normals" who possess a fair amount of ego strength.

Rostker (103) suggested, on the other hand, that personality is not a relevant factor in teaching ability but that the measured intelligence of the teacher is the highest single factor. Knowledge of subject matter and the ability to diagnose and correct pupil scores were found to be statistically nonsignificantly related to designated teaching ability.

Fielstra (38) isolated twelve characteristics of first-year secondary school teachers which correlated positively with principals' ratings. The characteristic discriminating most between teachers rated "good" and "excellent" was adaptability to a variety of teaching situations. This finding was similar to those resulting from studies with junior college teaching interns which will be reported in a subsequent chapter.

Conversely, Ort (83) suggested that neither academic achievement in college nor results of personality, attitude, or other tests seemed to have significant value for predicting how successful a student would be as a student or first-year teacher. He also pointed out the difficulty in controlling the many variables involved in teacher evaluation. Drive, motivation, level of students, philosophy, health, experimental backgrounds—or a combination of these—may all be important determinants of personality but are difficult to isolate. However:

> The best predictions of the future success of a student teacher, even though limited, can be made by the supervising teacher and the campus supervisors. The narrative description (which was) made by the supervising teacher and the campus supervisor concerning the student teacher, together with the scale evaluations made by these supervisory persons, provides the most valuable type of recommendation (83:70).

SUMMARY

This chapter has reported only a very small sample of the research concerned with individual personality characteristics as they relate to, and may be predictive of, teaching effectiveness. The literature is vast, the approaches are varied, and the techniques are diffuse. The answers, however, are far from determined. The question of criteria is a particularly important one and accounts for much of the disparity that exists between numbers of studies and the extent to which they are consistently implemented in practice. The pleas for more research are valid but replicable measures must also be considered. And, research on teaching effectiveness needs not only specification of criteria of effectiveness but careful definition of goals and objectives upon which to base the independent variables--the results of pupil-teacher interactions in educational situations.

RATINGS OF JUNIOR COLLEGE TEACHING INTERNS

chapter III

PERFORMANCE PREDICTION

The prediction of performance continues to be a central issue in educational research. It calls forth a variety of measures and a host of approaches, all of which seem to raise more questions than they answer. Differences in population samples, educational levels, and media are apparent in much of the research on teacher personalities. More striking, however, is the great variation in terms of theory—and often, the absence of theory—upon which research designs are structured. With so much emphasis placed upon this kind of investigation, it is interesting to note that there is yet no theory of personality upon which to build a discussion of the effective teacher; indeed, many theories may prove to be relevant. Still, there appears to be a need for theory or, at least, a set of assumptions that can be tested in these many efforts to relate teaching effectiveness to teaching performance and both variables to the personality constellations of involved individuals.

That many of the research findings are inconclusive and that research results have not been directly applied to contemporary educational situations may well make us suspect. Why aren't the results of these many efforts implemented in practice? What theories of personality would apply to teachers as they function in their many roles? Is there actually a type or set of characteristics that may be isolated? Or must we find new ways to define teaching effectiveness and proceed from that point to developing new theory upon which predictions can be made?

Does the research at one level of education apply to other levels?

Even today, with higher education such a force in our society, much of the research is concentrated on elementary and secondary levels education. Because people are people, does this mean that results of studies with kindergarten teachers are relevant to junior college instructors or university faculties? Do junior college instructors exhibit heterogeneity to a degree that students are exposed to a variety of types of adults? Or are the men and women who teach in community colleges similar to each other in academic backgrounds and previous work experiences? Which of the students who enter teacher preparation programs will be deemed successful by their university and junior college supervisors? Is it possible to identify personality characteristics that would lead certain people more easily than others to make the transition from student role to teacher role?

INVESTIGATIONS

Questions of this sort have stimulated a series of investigations with junior college teaching interns who have been enrolled in the Graduate School of Education, University of California, Los Angeles. Part of an ongoing research project, these studies are concerned with two major dimensions: the assessment of teachers as individuals, and the evaluation of overall teaching performance on the basis of specified criteria. Accordingly, they concentrate on certain questions that are often probed but seldom deal particularly with the junior college faculty. Further, they are based upon a theory of personality that views the individual in terms of dynamic forces and perceives the ego as a core dimension. It is not a theory that can be attributed to one person but is, rather, an eclectic approach to personality functioning, drawing heavily upon the work of such men as Freud, Jung, Murray, Erikson and Kris. Ego psychology—with its focus on "inborn ego apparatuses" dealing with intrapsychic economy and external reality—is still in its incipient stages of development. Yet, no contemporary theory of personality can ignore the realm of basic self with which it is concerned, nor disregard the dynamic forces that motivate it.

These concepts are just as important for the educator, the college administrator, the university researcher, and the junior college faculty member as they are for the more traditionally oriented psychologists. They are consistent with certain statements of Trent and Medsker (124) who point out that for the student:

> Education . . . must be concerned with human development as much as with training for specialized skills. It must assert the values of self-direction, creativity, and flexibility as firmly as the importance of readiness for a particular job (124:4).

RATIONALE AND PREMISES

The underlying premise of the U.C.L.A. investigations was that performance—whether it be teaching or behavior—is an outgrowth of the basic personality pattern of the individual and his interaction in a particular environmental setting at a particular time. In order to assess performance, then, there must exist understanding of the individual as well as very specific criteria (of which he is aware). Thus, a structured and defined basis for evaluation can be developed.

The person involved in changing his role from student to teacher is in a state of flux. He must adjust to many new demands. Considering this transitional, often difficult state, investigations of the interns were based on the specification of certain global dimensions: (1) whether there were, indeed, particular types of individuals who could be considered "good" and "poor" teachers; (2) the general "adjustability" of these individuals and their ability to endure the move from one situation to another; (3) the degree of adaptive-flexibility which the "good" intern manifested as compared with that demonstrated by interns who were judged less effective in their subsequent teaching positions. Point three was of special concern and was strongly emphasized in the investigations. However, other characteristics were also examined in the attempt to discover which of the many variables of human personality might best correlate with future success in teaching.

Assuming, then, that the ability to make a successful transition—from student participant in the intern program to beginning teacher in a new environmental setting—is a variable characteristic, a prime focus of the investigation was resolution of the question: Is it possible to determine in advance of actual placement what types of people possess the necessary ability to adapt to new situations—especially as these are typified by junior college teaching? A basic hypothesis explored in these investigations was that ego strength, as measured by a particular projective technique, would relate to the success of neophyte teachers as appraised independently by their supervisors.

Results of the investigations of teacher types have already been reported in an earlier monograph in this series. They will not be repeated here except to note that:

> . . . subjects indicating preferences high in the feeling dimension (according to Jungian typology) were more likely to be employed in first-time teaching positions and that, after several months as junior college instructors, they were given higher ratings by their supervisors than were those subjects who had demonstrated preferences for the thinking dimension. . . . Other findings suggested that no one type of person is employed as a first-time teacher in the junior college to the exclusion of other types and that no one type of individual teaching as a first-time instructor in a junior college is rated higher than other types . . . the subjects failed to cluster in a single type of group or groups. Thus, the heterogeneity of the student

population in the junior college would seem to be matched by the heterogeneity of first-time teachers and teaching applicants (15).

The basic premises for these projects were that:

1. The degree to which a subject is able to integrate unstructured material will provide clues as to how he will build his courses.
2. The general adjustment level of an individual may be found in his responses to ambiguous stimuli.
3. The degree of adjustability that may be identified from certain designated responses is related to the degree of adjustability the subject will manifest in other unstructured situations and to other ambiguous events.
4. Ego-strength ratings will be related to an individual's ability to move from one type of situation to another in a particular manner.

The Adaptive-Flexibility Inventory was designed for the specific purpose of assessing ego functioning. Several premises were also basic to the formulation of this instrument:

1. Ego strength is an essential ingredient of the individual personality.
2. In order to understand the concept of ego strength, there must be a consistent, operational definition of the term.
3. Ego strength manifests itself in several ways which are reflected in individual behavior and which may be measured.
4. Behavioral patterns may be predicted by evaluating the several dimensions postulated as emanating from this core personality feature.
5. Ego strength may be assessed by means of a projective technique, specifically a word-association scale.

A procedure of analysis was developed which examined correlations of ratings on the Rorschach Technique (100), the Myers-Briggs Type Indicator (80), and the Adaptive-Flexibility Inventory (14), together with ratings by independent supervisors. This analysis was based upon the preceding premises as well as upon the assumptions that:

1. There is a definite interplay between the individual, the demands of his vocational situation, and his environment.
2. The degree of adjustment to environmental forces faced by the individual is related to the degree of congruence between his needs and the external environmental press.

Following these premises and assumptions, specific hypotheses were developed and tested:

1. Persons in the broad middle-range of ego strength will be judged to be successful as first-year junior college teachers.
2. Persons of very low ego-strength ratings will be deemed unsuccessful as first-year junior college teachers.
3. Persons with extremely high ratings of ego strength either will

be similarly judged unsuccessful or will not, by their own choices, remain in junior college teaching positions.

INSTRUMENTS

Several general assumptions were basic in the selection of psychological instruments for these studies and in relating their assessments to supervisors' rating. The primary purpose for selecting, evaluating, and correlating was to attempt to predict potential success in teaching. To this end, rationale was examined and premises and hypotheses were developed.

The choice of instruments was based upon the individuals concerned, the nature of the colleges in which they would work, and the demands of the tasks to which they would have to react. While the particular occupational settings were not measured, the general nature of the community college and the variety of students it serves was considered.

The battery of psychological instruments was administered to the intern groups early in their summer training program and prior to their assumption of teaching responsibilities. The instruments employed were:

1. The Rorschach Technique, administered in group form according to Harrower (54) but also including a ten- to twenty-five minute inquiry conducted individually with each subject. Here questions regarding determinants, locations, and popular responses were clarified.
2. The Myers-Briggs Type Indicator (80), given to both the intern groups and the nineteen subjects in 1966-1967 who had applied for internships but had not been selected to teach in the junior colleges. These people were called "candidates" rather than "interns."
3. The Adaptive-Flexibility Inventory, Form B-2, administered in group form.

THE RORSCHACH

The Rorschach, a technique to assess behavior by sampling the manner in which individuals perceive unstructured material, is surrounded by theory and hypotheses too extensive to be reviewed here (100). Briefly, it is a psychological instrument that consists of ten ink blots of varying designs and colors. These are shown to the subject, one at a time, with the request that he respond to "What does it look like to you?" Klopfer defines the technique as providing a

> . . . relatively standardized situation in which behavior can be observed. The assumption is that, on the basis of this limited sample of behavior, it will be possible to predict other kinds of behavior on the part of the subject in other situations. As a projective technique, the Rorschach has the further characteristic of providing a relatively ambiguous stimulus situation which will enable the subject to optimally reveal his individuality of functioning. (64).

When scored and interpreted by trained professionals, responses to the blots furnish a multidimensional description of the dymanic forces of the subject's personality and the manner in which he reacts both to environmental influences and to his own inner promptings.

The Rorschach protocols for the junior college teaching interns were scored according to the Klopfer system. Each was assessed on the basis of four approaches: (1) a quantitative or sign assessment; (2) the Rorschach Prognostic Rating Scale; (3) a global rating of general adjustability; and (4) a global assessment of cognitive-integrative level.

1. The sum of ratings of eleven specific signs was employed as a measure of general personality adjustment. In selecting these particular signs, the interpreter was aware of the fact that these subjects would soon be involved in teaching in junior colleges and that the environmental presses they would encounter demanded a certain amount of flexibility and adjustability.

Three of the selected signs were attributed ratings on a five-point scale; eight others were assigned ratings on a three-point scale, thus providing a total possibility of thirty-six points of general adjustability. The Rorschach signs were: maximum, minimum, and average-form level ratings; movement ratings; F per cent; movement and color relationships; FK, Fc, and F relationships; the ratio of FC: (CF+C); A per cent; and chromatic: achromatic ratios.

2. The Rorschach Prognostic Rating Scale (RPRS) was developed as a measure of ego strength and as a predictor of response to psychotherapy. For purposes of these investigations, this scale was used because ego strength is considered to be a core dimension in personality functioning and, therefore, an important variable by which to evaluate the intern group.

Designed to quantify in an objective way the "intuitions" or "hunches" of experienced clinicians, the RPRS scheme is designed to rate each response for form level and for five Rorschach determinate categories. These are assigned differential quantitative weights according to a specific set of qualitative criteria. Klopfer describes the scale as measuring "the adjustment potential of the individual" and suggests that its various sections are intended to

> differentiate the concept of ego strength in its most important components: reality testing, emotional integration, self-realization, and mastery of reality situation . . . (It taps) the combined total of (1) the adjustment level or available ego-strength . . . and (2) the unused portion of the developmental and adjustment potential (64).

3. Each Rorschach protocol was given a global rating, ranging from one to five, based upon clinical "intuition" that reflected the interpreter's estimate of the subject's tendency to integrate his experiences. Since the Rorschach technique samples the way in which individuals react to new and unstructured situations, the rating was in-

tended to serve as a global assessment of each intern's potential capacity to integrate the demands of the new situation which would soon confront him as a junior college teacher.

4. A second global approach was used to assess the integrative level of each Rorschach protocol. In this respect, the interplay of determinants was considered to establish a general rather than strictly numerical system for evaluation. High's and low's in form level, number and quality of M responses, proportion of pure form determinates, number of responses, the types and per cent of location categories, and the amount of differentiation within the W (whole) responses were all considered.

On the basis of these criteria, the Rorschach protocols were rated on a five-point scale which followed the first five levels assigned by Bloom in the *Taxonomy of Educational Objectives:* 1=Knowledge; 2=Comprehension; 3=Application; 4=Analysis; and 5=Synthesis (12). If, for example, a subject were judged to be functioning at the level of Application, he was given a rating of "three." If he seemed to meet the criteria for both the Application and Analysis groups, his rating would be (3+4) 3.5.

MYERS-BRIGGS TYPE INDICATOR

One of the several attempts to classify human beings according to psychological types, the Myers-Briggs Type Indicator, is a self-administering technique based upon the conceptual scheme devised by Jung. This use of this instrument with the interns and teaching candidates have been reported elsewhere (15).

ADAPTIVE-FLEXIBILITY INVENTORY

The *Adaptive-Flexibility Inventory* is a word-association list, developed on the rationale that ego strength is

> . . . a concept which refers to the various functions of the ego in its relationship to both outside reality and to the larger Self. Ego strength may be demonstrated in the degree of adaptive-flexibility which an individual possesses. It represents a composite of dimensions, any or all of which are present within the individual to varying degrees. Together, these dimensions comprise the overall area designated as ego-functioning:
>
> 1. The ability to rebound, to emerge from challenging experiences
> 2. The ability to delay gratification
> 3. Toleration of ambiguity and conflicting forces, both internal and external
> 4. Acceptance of complexity
> 5. Flexibility rather than constriction and/or authoritarianism
> 6. Energy and creativity
> 7. Intelligence
> 8. Good reality testing
> 9. Sufficient experiences to provide the ego with opportunities to gain strength through growth

10. Ability to relate to the unconscious, to become subservient to the Self, and to tolerate regression when necessary for greater development, to meet the demands of the Self. This is at the highest level of development (14).

The Adaptive-Flexibility Inventory is presently used as a research tool to evaluate the degree to which an individual functions according to various specified dimensions of the basic ego-strength construct. These qualities are considered relevant to success in meeting the demands of changing environments and thus, to success both in the teacher preparation program and in the first year of teaching.

Responses to the word list are evaluated in a global or holistic manner on the basis of a seven-point scale. Low ratings describe people of low ego strength who will have difficulty fitting into certain types of environments (for example, the junior college) because they lack sufficient flexibility, creativity, and adaptability to meet the demands of the situation. People at the upper limits of ego strength (six and seven) are seen as demonstrating extremely high ego functioning but not fitting into certain situations because they have too much flexibility to feel compelled to stay in situations which they may perceive as limiting.

SUBJECTS

Candidates for U.C.L.A. Junior College Teaching Internships, 1964 through 1967, were recruited through word of mouth, through the Educational Placement Office, and by means of direct response to posters which had been distributed on the campus (25). In all cases, the candidates might be said to have selected themselves.

Each applicant to the program was interviewed by representatives of the placement office as well as by the director of the internship program and his staff. Those who appeared to be unlikely teaching prospects (because, for example, of blatant personality problems or extreme physical disabilities which would render their attaining positions extremely unlikely) and those who failed to meet intellectual and academic entrance requirements were eliminated. Others, considered to be likely prospects, were given information about specific junior colleges that were seeking teachers in their fields of concentration. These men and women worked directly with the placement office and the community colleges to secure positions as teachers for the following academic year, a procedure which also served as a further screening device.

From the more than 120 applicants who initially sought positions as junior college interns, 46 were selected in the three-year period with which this report is concerned. There were 20 interns in the 1965-66 program. Nineteen of this group were later assessed by both the program director and the college supervisors, one having dropped before the summer course was completed. In the 1966-67 program, group of subjects was based upon their responses to the Myers-Briggs Type Indicator and the Adaptive-Flexibility Inventory while the data concerning the interns included responses to all three instruments plus the independent supervisor judgments.

Table I presents ratings assigned to all thirty-seven interns who had engaged in the testing program. For the Rorschach, ratings are based upon the four specific approaches discussed earlier: total of the thirteen signs, the Rorschach Prognostic Rating Scale, the global rating of general adjustability, and the global assessment of cognitive-integrative levels. Interpretation of the Adaptive-Flexibility responses are handled as a single global measure ranging from one to seven.

The Myers-Briggs preferences indicated by the unsuccessful candidates in 1966-1967 and the intern groups of 1965-1966 and 1966-1967 were tabulated separately. Table II presents the number of subjects whose responses fall into the various type of categories designated by this inventory.

And finally, various ratings assigned to the 37 interns of 1965 through 1967, on the bases of their responses to the psychological instruments have been totalled in Table III. For the Rorschach ratings, the greatest numbers are found in the top group for the total sign approach; the middle group for the RPRS; the third and fourth groups (of five) for the global assessment of general adjustability; in the fourth group for Bloom-level ratings; and, for the total Rorschach evaluation, in the fourth of five groups.

The Adaptive-Flexibility ratings of ego strength suggest that nearly half the interns (fourteen of thirty-seven) fall into group five, the highest category of the average group. Here there are no "sevens" and but one "one"—a subject who was later found to be the only person in all the intern groups unable to complete even the summer pre-service program. Speaking for environmental press/individual functioning relationship, basic to assessment on the A-F scale, the many interns falling into the high-average group are those who are seen as most likely to succeed in junior college teaching roles.

Table IV presents ratings assigned to each of the interns (N=34) completing the program by the U.C.L.A. director and his college supervisor.

Correlations initially computed for the 1965-1966 interns on the basis of selected Rorschach signs showed that seven of the thirteen signs were positively related to supervisor's ratings.

Since the thirteen signs are of greatest meaning when considered as a whole, it was decided to correlate only the total of the 13 signs for the combined responses of all participating interns. However, the data may be of general interest and are presented here for that reason.

Table VI presents an intercorrelation matrix for the four Roschach

17 subjects were assessed by psychological instruments and 14 were rated by their college supervisors and the program director. Another group was also included in the investigation for this same year—individuals who had applied for membership in the program but had not been selected for teaching positions by the junior college placement officers. These applicants, not successful in becoming interns, were designated as "candidates." Together they included nineteen people.

PROCEDURE

Each candidate who had been successful in obtaining a junior college position was tested during the summer preceding his entry into teaching. The tests were scored by the junior author and by a graduate student in clinical psychology; ratings on the A-F Inventory and Rorschach interpretative scores were also assigned. Each intern was rated by two independent judges:

1. The U.C.L.A. program director, the senior author, rated each of his students on a 5-point scale ranging from poor (1) to good (5). Criteria for his assessments were the subject's tendency to integrate the preparation sequence demands, as demonstrated by construction of course outlines, and the intern's submission of evidence that his students had learned under his direction.

2. The junior college supervisors (deans of instruction or department chairmen, as determined by each school) judged each intern—the first-year teacher—on a five-point scale with one suggesting a low rating and five a high rating. Data were collected by independent assistants in the program who asked for (1) the supervisor's global evaluation of each individual's teaching performance, and (2) his relationship with faculty, administrators, and the community at large. The supervisors were also directly asked: "If you had known the intern then, at the initial time of hiring, as you know him now, would you have employed him?" The program director, the junior college supervisors, and the psychological instrument evaluator all acted independently. None was aware of the others' ratings. The director's and the supervisors' ratings constituted the independent criteria of ability to make an adequate adjustment in the changing role of student to teacher.

RESULTS

Responses by the interns to the Rorschach, the Myers-Briggs, and the Adaptive-Flexibility scales were all assessed, according to criteria previously described, for three general purposes: (1) to understand the personality characteristics of the interns; (2) to discover any patterns or specific types that might become apparent; and (3) to correlate the independent criteria (judgments of the program director and the college supervisors) with the test material. The data were also analyzed in three parts: (1) the 1965-1966 group of twenty interns; (2) the combined group of thirty-seven interns of 1965-1966 and 1966-1967; and (3) the thirty-seven interns and nineteen unsuccessful candidates of 1966-1967. Information regarding this third

Table I:
PSYCHOLOGICAL INSTRUMENT EVALUATIONS FOR INDIVIDUAL SUBJECTS

(Note: 100 series = 1965-1966 interns; 300 series = 1966-1967 interns)

	The Rorschach					
Subject	Total of 13 Signs	R.P.R.S.	Global Adjustment	Global Bloom Level	Total of Rorschach Ratings	The A-F Scale
101	27	8	3	3.5	41.5	1
102	27	13	3	4.5	48.5	3
103	34	8	5	5	52	5
104	31	13	4	4.5	52.5	6
105	31	9	3	4	47	5
106	22	1	3	2.5	28.5	3
107	30	10	3	3	46	4
108	25	8	3	4	40	6
109	33	16	4	5	58	5
110	28	9	2	4.5	43.5	5
111	32	11	4	5	52	4
112	31	11	3	3	48	5
113	33	8	4	4	49	4
114	31	11	5	5	52	3
115	26	8	3	2	39	4
116	29	13	3	5	50	5
117	32	15	4	4	55	5
118	25	3	2	1.5	31.5	2
119	31	16	5	5	57	6
120	28	13	3	3.5	47.5	5
301	29	11	4	4.5	48.5	5
302	27	5	4	3	39	5
303	32	12	4	4.5	52.5	4
304	25	7	3	3.5	38.5	4
305	26	13	4	3.5	46.5	6
306	Did not take Rorschach					2
307	26	9	4	4	43	6
308	31	18	5	5	59	5
309	29	15	4	4	52	4
310	33	14	4	5	56	4
311	29	4	3	3	39	5
312	22	4	3	3	32	3
313	32	10	4	4	50	5
314	33	14	4	4.5	55.5	3
315	26	12	4	4.5	46.5	4
316	31	9	3	3.5	46.5	5
317	29	15	3	3.5	50.5	6
N = 37						

Table II: NUMBER OF INTERNS AND CANDIDATES DESIGNATING SPECIFIC TYPE PREFERENCES

	No. Interns, 1965-1966, Designating this Preference	No. Interns, 1966-1967, Designating this Preference	No. Candidates, 1966-1967, Designating this Preference	Total
(E) Extravert	17	7	11	35
(I) Introvert	3	10	8	21
(T) Thinking	7	5	11	23
(F) Feeling	13	12	8	33
(S) Sensation	6	4	4	14
(N) Intuition	12	13	15	40

N = 20 N = 17 N = 19

Total N = 56

Table III:
NUMBER OF INTERNS FITTING INTO VARIOUS PSYCHOLOGICAL INSTRUMENT RATING CATEGORIES

		1965-1966 Interns	1966-1967 Interns	Total Number
Rorschach Ratings				
Total Sign	34-30	11	6	17
	29-26	6	6	12
	25-22	3	4	7
R.P.R.S.	18-13	7	6	13
	12-7	11	7	18
	6-1	2	3	5
Global Adjustment	5	3	1	4
	4	6	10	16
	3	9	5	14
	2	2	0	2
	1	0	0	0
Bloom Level	5	6	2	8
	4.5-4	7	7	14
	3.5-3	7	7	14
	2.5-2	2		2
	1.5-1	1		1
Total of Rorschach Ratings	53-59	3	3	6
	46-52	11	8	19
	39-45	4	2	6
	32-38	0	3	3
	25-31	2	0	2
Myers-Briggs Ratings	E	17	7	24
	I	3	10	13
	T	7	5	12
	F	13	12	25
	S	6	4	10
	N	14	13	27
Adaptive Flexibility Ratings	7			
	6	3	2	5
	5	8	6	14
	4	4	5	9
	3	3	2	5
	2	1	1	2
	1	1	0	1

Table IV:
INDIVIDUAL INTERN RATING, INDEPENDENT SUPERVISORS

Subject	SUPERVISORS	
	U.C.L.A. Director	College Supervisor
101	1	dropped
102	3	2
103	2	4
104	3	3
105	2	3
106	1	1
107	1	2
108	2	4
109	3	4
110	3	4
111	2	4
112	3	4
113	2	2
114	3	3
115	1	2
116	1	2
117	2	4
118	1	1
119	3	4
120	3	2
301	5	5
302	3	5
303	4	4
304	2	5
305	2	5
306	dropped	dropped
307	2	3
308	2	3
309	3	5
310	3	5
311	4	4
312	dropped	dropped
313	5	5
314	2	4
315	4	5
316	4	5
317	4	4

Table V:

CORRELATION OF INDIVIDUAL RORSCHACH SIGNS, AND TOTAL OF THE 13 INDIVIDUAL SIGNS WITH SUPERVISOR'S RATINGS

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Maximum Form Level Rating	03	55	−42	21	29	43	02	43	15	09	43	40	61	36	49
2. Minimum Form Level Rating		36	−24	−16	29	−02	−06	−24	−39	70	−40	−42	12	−08	−02
3. Average Form Level Rating			−47	45	31	22	10	−05	−03	36	−18	−06	45	43	40
4. F^2				−07	−36	−30	−05	04	10	−40	23	11	−07	−23	−43
5. M					18	39	24	32	55	−10	19	52	49	57	42
6. FM						79	−23	29	−07	74	−02	18	39	16	53
7. FM + m							−40	53	09	37	10	42	46	28	59
8. FK + Fc								−15	28	−11	06	08	03	25	−10
9. FC									35	−06	64	85	50	26	45
10. CF										−38	50	73	22	36	31
11. A^2											−40	−26	10	00	27
12. (Fc + c + C^1)												−71	36	08	24
13. (FC + CF + C)													55	39	45
14. Total of All Signs														60	49
15. Junior College Supervisor Ratings															67
16. U.C.L.A. Program Director Ratings															

Decimals omitted

$r = .38$ $p < .10$ $r = .44$ $p < .05$ $r = .56$ $p < .01$

Table VI:
CORRELATIONS OF GLOBAL AND QUANTITATIVE RORSCHACH RATINGS AND SUPERVISORS' EVALUATIONS FOR INTERNS

	1	2	3	4	5	6
1. Total Rorschach Signs (selected 13)		.62	.51	.63	.60	.49
2. Rorschach Global Rating			.45	.65	.41	.39
3. RPRS (total of six scales)				.44	.39	.35
4. Rorschach Integrative Level					.60	.54
5. Junior College Supervisor Ratings						.67
6. U.C.L.A. Program Director Ratings						

$r = .38 = p < .10$
$r = .44 = p < .05$
$r = .56 = p < .01$

ratings and the supervisors' judgments for the 1965-1966 interns effectiveness as beginning teachers.

The four approaches to interpretation of Rorschach data all correlated positively with each other. Two of these ratings (total of thirteen selected signs and the Rorschach global assessment of integrative ability) were then tested against the Adaptive-Flexibility ratings and the evaluations by both the U.C.L.A. program director and the college supervisors. These revealed the relationships for the 1965-1966 interns as shown in Table VII.

Table VII:
CORRELATIONS OF TWO RORSCHACH RATINGS AND SUPERVISOR RATINGS FOR 1965-1966 INTERNS

	1	2	3	4
1. Adaptive-Flexibility Inventory				
2. Rorschach Sign Approach	.31			
3. Rorschach Global Assessment	.29	.59**		
4. U.C.L.A. Program Director	.45*	.44*	.54*	
5. Junior College Supervisors	.56**	.57**	.61***	.68***

* $p < .05$
** $p < .01$
*** $p < .005$

Table VIII: CORRELATION OF INSTRUMENT ASSESSMENT, PROGRAM DIRECTOR'S RATINGS, AND SUPERVISOR'S RATINGS OF JUNIOR COLLEGE BEGINNING TEACHERS

	1	2	3	4	5	6	7	8
1. Adaptive-Flexibility Inventory (7-point scale) Global Rating (7-point scale)		.23	.20	.26	.36	.27	.36	.39
2. Total Rorschach Sign Approach			.50	.48	.56	.60	.29	.26
3. Rorschach Global Adjustment Rating				.39	.52	.66	.23	.29
4. RPRS					.59	.45	.32	.28
5. Rorschach Raw Movement Scores						.67	.25	.28
6. Rorschach Integrative Level							.29	.34
7. U.C.L.A. Program Director								.63
8. Junior College Supervisor								

N = .34
r = .31 $p < .05$
r = .43 $p < .01$

For this group, the two Rorschach assessments show a strong, positive correlation (.57 and .61) with college supervisors' ratings. They correlate less well (.44 and .54) but positively with the ratings attributed to the interns by the program director. Similarly, the Adaptive-Flexibility Inventory ratings correlate positively (.56) with the in-service supervisors' ratings and, to a lesser degree (.45), with the ratings of the program director. The highest correlations obtained are those of the program director and the college supervisors (.68). This substantiates the contention that ratings made by university training supervisors consistently prove to be the best available yardsticks for predicting judgments made by supervisors in work situations (83).

Correlations were computed as the Pearson r for all junior college teaching interns who completed the program, a total of thirty-four of thirty-seven subjects (since three had dropped out before supervisors had an opportunity to rate them). These computations were based on the individual subjects ratings described in Tables I and III and are presented in Table VIII.

For this number of subjects, a correlation of .31 is significant at the .05 level and a correlation of .43 is significant at the .01 level. The Adaptive-Flexibility global assessments correlate with program director (.36) and supervisor's ratings (.39) better than any of the five Rorschach ratings, although the correlation of the RPRS and the independent ratings (.32) is positive and significant at the .05 level. The difference between correlations for the 20 interns of 1965-1966 alone and the combined groups may be attributed to the more limited spread of the ratings which were ascribed to the second-year interns. The distribution of ratings by supervisors for 1965-1966 interns and 1966-1967 interns is shown in Table IX.

Table IX:
DISTRIBUTION OF INTERN RATINGS BY SUPERVISOR'S

Rating	*No. of Interns*	*Rating*	*No. of Interns*
5	0	5	0
4	8	4	4
3	3	3	2
2	6	2	0
1	2	1	0
	19		15

DISCUSSION

Several questions were asked in the course of our intern appraisal, for example: "Is it possible to determine in advance of actual placement what types of people possess the ability to move successfully from student to teacher roles? Which students are able to make the necessary transition rapidly and well enough so that they will be deemed successful in their initial teaching experiences?"

Ratings on the Adaptive-Flexibility Inventory positively and strongly correlated with independent ratings of teaching success for the 1965-66 interns. When these same subjects were evaluated together with junior college teaching interns of the following year, 1966-67, the correlations were lower but still positive. This difference poses an interesting problem as to why predictions for one year differ so greatly from predictions by the same assessor in a subsequent year. One possible explanation for the inconsistency is that the second group of interns tested were better prepared by the U.C.L.A. program than was the previous group. Therefore, although the *A-F* scale may well have picked up the same types of personality interpretation, the abilities to cope with new teaching situations might have been better developed.

A second possible explanation is that the criterion variable, supervisor's ratings, is itself somewhat unreliable and may account for some of the failure to obtain strong positive correlations with the combined groups of interns for the two successive years. The third-year interns (1966-67) were generally rated higher by college supervisors than the second-year interns, perhaps owing to the prestige of the program ("halo effect") or to the better selection of third-year interns. The ability to get meaningful data to establish the predictive validity of the Adaptive-Flexibility scores was substantially reduced because of the high supervisor ratings. This restriction in range makes high correlations statistically improbable. There is a need to find some way of forcing a greater spread in supervisor's ratings for future studies if higher correlations are to be expected.

Although none of the hypotheses relevant to the Myers-Briggs Type Indicator was supported beyond a low degree of correlation, some tendencies toward specific typology did appear. These have been reported in an earlier monograph in this series.

SUMMARY

From these somewhat limited findings concerning U.C.L.A. teaching interns, there would seem to be a variety of types of people who enter junior college teaching. This mix of types suggests at least a partial reply to the question of whether current practices of teacher selection in colleges allow youth to gain an appropriate range of teacher types who can serve as models. On the basis of demographic and psychological data collected on this group of new teachers, it would seem they do. The question of whether instructors exhibit variety in their situational behavior was not investigated as part of this study but might be inferred if we assume that these measured characteristics are basic and reliable indices of underlying traits.

These studies have been reported in some detail because they stem from theory—hence have potential for advancing knowledge of human functioning—and because they offer a way of viewing beginning teachers with a fair chance of selecting those who will be adjudged successful. They still do not get at the question of ultimate criteria, however, because they interpret "success" as "achieving a high rating from a supervisor." Ways of measuring success as it relates to learning attained by pupils will be examined in the next chapter.

PART II

SCENE: A HARDWARE STORE

CLERK: May I help you?

CUSTOMER: I'd like to buy a tool.

CLERK: Yes, sir. What kind?

CUSTOMER: Something to measure with.

CLERK: Yes. What do you want to measure?

CUSTOMER: I'm not sure.

CLERK: Well, what kind of tool did you have in mind?

CUSTOMER: Oh, the kind of thing my friends use. A micrometer; perhaps a yardstick. One fellow I know uses a surveyor's level. One of those maybe.

CLERK: It would really help if you could tell me what you intend to measure.

CUSTOMER: I don't know. I can't define it.

CLERK: Well, why do you want to measure it?

CUSTOMER: Can't say, but other people have been measuring it for years.

A CRITIQUE OF CURRENT PRACTICES

chapter IV

Generally, difficulties in assessing instructors can be traced to two sources—ambiguity of purpose and indeterminate criteria. Until those issues are resolved, all rating schemes are doomed to severe and legitimate criticism, if not to abject failure.

DIFFICULTIES

Studies of faculty appraisal at all levels of education rarely examine in depth the reasons for evaluation, in spite of the fact that purpose must be at the core of all schemes. Only after questions regarding the purposes for examination have been carefully studied can an institution settle upon a scheme which will be at least reasonably satisfactory to all parties. If we say that teaching can be evaluated, we assume there exists an acceptable definition of teaching. If the definition of teaching is accepted as "causing learning," we assume that learning can be appraised in some objective fashion. Thus, the institution that accepts the definition of *teaching* as *causing learning* has taken an important step toward bringing order into its faculty evaluation processes.

The criteria, then, for each evaluation must become the demonstrations of student learning presumed to have resulted from the efforts of the teacher in question. However, although learning achieved by students can be used as a major criterion in assessing faculty members, it is not a sufficient condition for devising a faculty evaluation scheme. Related to the issue of criterion designation is the concommitant matter of purpose for evaluating faculty upon the selected criteria.

REASONS FOR EVALUATING

In spite of the popularity of evaluation schemes, the purposes for measuring instructors in the junior college are nebulous and the reasons for continuing such practices are often unrelated to criteria. It is, in fact, difficult to differentiate among real and apparent reasons. The reason for appraisal is often said to be "to improve instruction," but the methods seldom relate to instructional practices and even less often to the results of instruction. As typically conducted, faculty evaluation cannot be seen as a way to improve instruction.

Another reason for evaluation procedures may be the administrators' tendencies to stand in judgment of instructors, just as instructors seem to stand in judgment of students by applying grade marks to them. Again, it may be the desire to say plaintively to the world, "we are good." Such a rationale has been applied in numerous articles on junior college teaching, including those that loudly proclaim "faculty competence" (11).

Simple inertia may be a further reason for appraising faculty in the community college. Most public school districts evaluate their teachers; many four-year institutions engage in similar practices. Why shouldn't the community college join the parade. It is frequently easier to continue a practice than to justify a change.

While the reasons for conducting faculty evaluation may not be identified with either purposes or criteria, some of the blame for the strange appraisal procedures must be ascribed to the nature of the profession itself. Despite teachers' longing to be "professional," they are not oriented toward examining the results of their efforts to cause changes in their clientele. Intentions are often mistaken for the realities of life in the schools. Further, there is much of a whistling-in-the-dark attitude of, "We are one big happy family of dedicated, competent professionals." Faculty evaluation schemes frequently serve to perpetuate that self-delusion; accordingly, when evaluation is tied to merit pay or tenure recommendations, it is occasion for despair. If administrators and instructors are, in truth, one large "family of professionals," an attitude of betrayal is fostered by so relating judgments to tangible rewards.

If purposes, criteria, and reasons for conducting appraisal are ambiguous, why, then, bother to evaluate at all? Is the purpose to allow administrators to stand in judgment of their faculty? If so, little wonder that instructors categorically reject the process! Development of a profession cannot be enhanced, they argue, if members of the profession are subject to external judgment because, by definition, a profession is self-policing. A corollary argument against judging faculty is that administrators, by virtue of their being removed from the teaching situation itself, are not qualified to assess teachers. Both contentions may have merit. However, if teaching is to be truly a profession, it must begin to police itself in order to counter external judgment.

Is evaluation to be used for the purpose of awarding merit pay? Nothing can cause as much uproar among a faculty as that suggestion! Statements made by three authors in a recent edition of *The Educational Forum* demonstrate how the lines are drawn:

> Whatever the theoretical virtues of merit pay may be, the special conditions of our profession deny justice in its application. Consider first the impossibility of applying an objective yardstick to the creative process of teaching. It cannot be measured as a salesman's sales for a given period can be measured, nor weighed as the amount of sand a hod-carrier shovels in one day can be weighed, nor even calculated as box-office receipts for an opera can be calculated (122).

The argument is that because it *cannot* be measured, it *should not* be measured. Conversely:

> Unfortunately, the failure of the education profession to advance beyond the landmark of an equitable salary schedule to the provision of widely accepted criteria and methods of rewarding good teachers for their excellence no doubt contributes to the shortage of talented people who become and remain classroom teachers (88).

And:

> If new methods of evaluating teacher effort must be created, let's get on with the task of creating them rather than continuing to perpetuate excuses for not doing so (91).

On the one side, then, are those who would say that teaching, as presently defined, cannot be measured. On the other are those who argue that unless provision is made for differentially rewarding teachers, the profession is doomed to mediocrity. The present system, they say, attracts a conforming, security-bent type of person to work within it.

Is evaluation to be used to determine which instructors shall be given continuing contracts? In some states—California, for example—an instructor, once employed, cannot be dismissed unless cause is proved. Interpretations of specific causes for dismissal are not clear but they include such matters as incompetence and moral turpitude. Gathering evidence on which to dismiss instructors is an unfortunate reason for introducing and maintaining a faculty evaluation scheme. It puts the evaluators in the position of doing the detective work—a task performed much more efficiently by an agency such as Pinkerton!

Shall evaluation be used to build a case for the worth of junior college instructors? Who must be convinced? The university will believe that junior college faculties are competent if students who transfer are well-prepared. As typically practiced, however, evaluation is far from enhancing student learning—no one has brought evidence to show that students learn less in institutions where faculty evaluation is not practiced. The public can better be convinced of the worth of a junior college faculty by a skilled public relations officer who writes a variety of glowing tributes than by reports written by evaluators. If the purpose for appraisal is only to convince one's colleagues and one's own self of competence, the practice should best remain haphazard and voluntary, awaiting each instructor's need for a particular type of approbation.

VALID PURPOSES

In spite of the common disparity between purposes and practices, there are many valid purposes for employing faculty evaluation. These relate directly to overall institutional goals and point to reasonable criteria.

An evaluation scheme can be employed to direct faculty efforts even where it is not tied to pay raises or other forms of extrinsic reward. Evaluation instruments and procedures are powerful forces in determining what goes on within schools and classrooms (59). For example, if evidence of student learning is to be gathered, instructors are more likely to direct their efforts toward causing learning than they would be if views of their performance alone were accepted as the major criteria of teaching worth. For that reason, the practice of making sound or sight recordings of instructors' performances, whether in training or in actual teaching situations, seems to have little merit in directing their efforts toward causing learning. Rather, it tends to focus attention once again on "quality" of performance and encourages instructors to continue assuming a connection between performance and learning achievement.

If it is to be used for the improvement of instruction, faculty evaluation must be related to instruction as a discipline. Instruction cannot be measured by observation alone because it is a multidimensional concept, a process by means of which a student's environment is so shaped that meanings are easier to grasp. It is a way of ordering perceptions so that learning occurs; without measuring the resultant learning, there is no way of telling whether or not a student's environment and perceptions were so ordered. And by short-range observations of instructors, there is no way of viewing total sequences. If instruction itself is to be measured, the assumption must be made that students can learn and that there are such things as sequences which can help them organize their thoughts.

A measurement of instruction assumes more. It suggests that the purpose of the junior college is to provide instruction to the young, not merely to provide them with models of well-functioning adults. It puts a negative value on mere observation of instructors, implying that instruction can be measured independently of the person of the instructor (although continuing investigations may show the relationship of certain personality characteristics to instructor effectiveness). Learning is an internal process which can be shaped by external forces. The person of the instructor is a force—but only one force—in the total learning environment. If the instructor is to be observed or rated as a *portion* of a total learning environment, methods other than those typically in use must be employed. And, most important, *effects* of the instructional process must be included in the paradigm.

Faculty evaluation may eventually prove to enhance the development of instructional specialists. Currently a junior college instructor must be competent in all aspects of the instructional process. This means he must be a scintillating lecturer, a stimulating discussion leader, an omniscient setter of objectives, a warm-hearted counselor, a skilled media producer, and a careful writer of examination items. To expect all persons working in the schools to be thoroughly competent in all facets related to the discipline of instruction is to doom the profession to mediocrity. Specialization must be enhanced so that the institution may be staffed by a core of people who collectively, but not necessarily individually, display excellence in all matters relating to teaching.

Instructional specialization suggests team teaching of one type or another, a practice becoming widespread among institutions at all levels of education. The instructiional team may have one of its members write objectives, another give lectures, a third select and produce replicable media, and a fourth construct and continually analyze test items (23). Within the team, each member must pull his share of the load or he adversely affects his immediate colleagues. In such cases, they can apply necessary sanctions to cause him to change or to eliminate him from the team. Each teacher would learn to do what he can do best and would add something of value to the group. Evaluation would then become a process by means of which one's fellows would influence his activities (68). Even now, whether or not instructors work together in teams, evaluation of a type other than that typically practiced might encourage them to specialize.

Evaluation can have broader purposes, too. The practice of instruction differs little now from the manner in which it was conducted one hundred years ago. On many campuses, instructors still do everything as it was done by their 19th century counterparts, short of shaking down the coal stove. It is not stretching a point to conceive of a modern junior college as a collection of little red school houses—boxes of isolation—one for each instructor who operates within a self-contained classroom. To what extent is the long established right of privacy of the classroom used indiscriminately as a shield of academic freedom to block all possible approaches to change (40)? Evaluation can have, as an important relevant purpose, the breaking down of such isolation by creating situations in which faculty members communicate with each other regarding instructional processes. A design built on this kind of rationale might have instructors evaluating each other *on the basis of their teaching effectiveness alone*. Joint consultations and visiting—practices which occur now informally—could be used more widely. It would not be necessary to involve administrators in such schemes.

SUMMARY

In its current form, the reasons for faculty evaluation are invalid and the criteria upon which it is based are nebulous. It fails to differentiate between the teacher as a social being and the effects of his teaching. It serves no useful purpose. It is time to abandon it and replace it with something of value.

An institution operating under any philosophical tent can have its view of the people it selects and retains made more meaningful if studies of them are tied to deliberate purpose. The junior college might well replace faculty evaluation with research on inputs to learning. What makes a difference in students' learning? Class size? Varieties of teaching materials? Multimedia instruction? Learning is occurring as a result of many forces—a few known, more surmised, and possibly still more unknown. One influence is likely the teacher himself. Knowledge of interactional effects of teacher and students is at a primitive stage but the junior college can profitably help advance it. Here study of people and of their effects can come together.

Similarly, junior colleges should know more about people who occupy faculty positions. This is important for enhancing the profession and for adding to a general store of knowledge regarding human functioning. If teachers are to be viewed as people, evaluation—or whatever replaces it—would be better conducted if it were tied to theory and to the rationale that knowledge can be furthered by the study of humans.

THE ULTIMATE CRITERION

chapter V

Problems in identifying criteria on which to base predictions of teacher effectiveness, in measurement of teacher performance, and in evaluation continue to plague the field. Even after decades of work in the area, research can say little to questions of teaching in the junior college and to faculty measurement. Conversely, evaluation practices in the junior college say even less to researchers. Unless research on teachers and practices of faculty measurement, supervision, and evaluation in the junior college come together on common ground, improvement of instruction in the junior college—a self-styled "teaching institution"—must suffer.

This chapter discusses problems in specifying criteria for assessing teaching effectiveness. It presents reasons for using student achievement of learning objectives as the main criterion upon which studies of faculty and of instructional effect should be based. Designs for assessing instructors and a scheme for supervising instructions are presented in the last chapter.

CRITERION SELECTION

In this monograph, the problem of inadequate criterion variables employed in studies of teachers and teaching effects has been mentioned frequently. Criteria suffer from equivocal definitions—an ailment compounded by the use of nebulous terms in measuring devices. Unreliability of measures and difficulties in getting various groups to agree on what should be measured, whether or not it should be measured, and what instruments to employ add to the problems.

However, the core problem with establishing criteria is the fact that unless they are validated against ultimate purposes of the institution, they fall short of adding anything of worth to institutional functioning. In the junior college, for example, instructors are typically rated by deans or division chairmen. Skirting questions of supervisor bias and halo effects; issues of unreliability in observation; inconsistency in terminology; and ambiguity in rating forms, what is the ultimate criterion upon which the rater bases his observations? In many cases, it is his assessment of whether or not the instructor under observation is likely to bring discredit upon the junior college. The instructor is often rated according to the extent to which the observer feels he will "play the game" and "not rock the boat." Naturally, perceptions vary and what is considered adequate behavior in one institution may not be so construed in another.

In a larger sense, a supervisor who rates a faculty member on his perceived "goodness" is using institutional self-perpetuation as the ultimate criterion. The instructor who unduly upsets people in the community by dress, habits, or untoward beliefs overtly expressed is more than an embarrassment; he is a genuine threat to institutional survival. It is true that institutions must attend to self-perpetuation if they are to function and achieve their educational purposes. But by definition, assessment schemes that look primarily to institutional perpetuation see that as an end in itself. And by so doing, much is lost—especially a chance for junior colleges to employ practices which add to understanding of human functioning and to knowledge of the discipline of instruction.

TEACHER-STUDENT EFFECTS

If, in the junior college, practices of questionable value and limited effect can give way to genuine study of people working in instructional situations, nothing would be lost. Instead, much might be gained. The institution could become a full partner in higher education, serving as a laboratory operated for the purpose of advancing knowledge about instruction—the junior college's main function.

To be effectual in changing practices in education, research must be indigenous. University-based researchers can design studies and make recommendations; however, change directed to satisfying the peculiar needs of junior colleges must result from studies conducted within them. Failing that, changes will continue to occur as a result not of research findings, but because of rhetoric, political persuasion, professional prestige of their advocates, fads and fashion (61).

An institution that is dedicated to teaching—causing learning—should study instructors, students, and the learning process. It needs answers to questions such as:

- What kinds of students' achievements result from classes taught by different teachers?

- From what kinds of students are such achievements shown?
- How is achievement of different kinds of students related to different conditions and environments?
- What does the teacher do that is related to the various kinds of achievements demonstrated by students in various kinds of situations and environments?
- What kinds of teacher experiences and personality factors are directly related to the kind and quality of teacher behavior revealed in relation to students?
- What processes of selection and education develop teachers with personality factors and behavior patterns shown to be associated with effective teaching behavior (95)?

The entire field of education would benefit if such questions were studied. In addition, research would enhance the development of theories of human functioning and of instruction, both of which are at present inchoate.

Merging streams of study on people and on instruction relate to the junior college in a narrower context. Helping develop theory and adding to a general store of knowledge about instructional process are noble ends, but there are more immediate payoffs for junior colleges electing to engage in the activities. As examples:

1) Selection of instructors can be made more relevant to institutional needs and purposes if it is conducted within a framework of objectivity and rationality.

2) Junior colleges must soon begin taking a larger responsibility for preparing their own instructors; university-based teacher preparation programs are simply not proving adequate (43).

3) Assignments of instructors to particular types of students can be undertaken with more certainty if they are made on defined bases rather than on vaguely expressed preferences or political suasion.

4) Instructional specialization, a phenomenon clearly on the horizon, must be carefully studied from points of view of feasibility, economy, and effect on student learnings. Instructors as people and instructional effects are clearly part of that study (23).

5) And, most important, student retention and achievement will undoubtedly be affected as knowledge of instructional processes increases.

Theories of instruction have not yet been distinctly postulated; (18) indeed, one may be struck by the absence of theory upon which to base pedagogy. In their place is a body of principles, axioms, and untested assumptions sometimes grouped under a heading called "the art of teaching." Consequently, it is difficult, if not impossible, to assess teacher competence by viewing teacher behavior. Few *teacher* behaviors, either in or out of the classroom, can be related directly to student learning because there are few *instructional* principles which can be so related. For that reason above all, researchers have been frustrated for decades in their efforts to seek magical formulae that relate teacher behavior to effective instruction. Research on teacher competence has proven to be noncumulative because relevant factors in instruction have not been identified and related to student learning (101).

Research on instruction itself has suffered similarly over the decades. What in instruction relates to what in learning? The question does not exist in the abstract because, although instruction may be conducted by a medium other than a live person, learning must be achieved by people. And people—in this case students—differ along more dimensions than any one study or group of studies can control. Unfortunately, however, people are too often left out of instructional design models.

Research on instruction occupies a vast place in the education literature, increasing over the past fifteen years with the development of various replicable instructional media. Auto-instructional programing alone has stimulated study of instructional variables like nothing else known in education. Whether or not programing replaces instructors or changes instructors' roles is moot; the point is that it has led to a renaissance in the study of instruction as a discipline.

As an instructional form, programing has come in for severe criticism because of its seeming failure to account for people (3). People are a part of the schools. Instruction can be studied in the abstract—indeed, it must be so studied if instructional theories are to develop—but people, both as practitioners and recipients, must be related to the practice of instruction. Despite the fact that the classroom with forty chairs facing a teacher's podium may not be a desirable, let alone effective, means of conducting instruction, it does exist and must be so considered in studies of the instructional process.

Research on teachers as people—competent, functioning in particular situations, being prepared, etc.—has suffered for a corollary reason. For one, teachers are too often viewed apart from their effects on students (other people). For another, people are too often viewed without consideration of the environments in which they labor; in short, interactions between personal functioning and characteristics of their world (15). Most seriously lacking, however, are attempts to correlate variables of human functioning with instruction as a field of study.

Instruction is a discipline which includes a body of concepts and established practice. Schools are staffed with people playing a variety of roles. *Study of one cannot proceed effectually without parallel study of the other.* Overlaps between the two exist in practice and in concept. It is therefore fruitless to attempt to stabilize theories of instruction without relating them to effects on practices in the schools and on student learning. Correspondingly, study of teacher behavior, personality, and competence cannot produce credible results unless the discipline of instruction is considered. A merger between the two must be effected.

The junior college is a logical place to combine studies of people with research on instruction because it is committed to instruction to a degree not present in other segments of higher education and because it involves a varied and increasingly growing population. Too, lines of influence are shifting within it. Instructors are demanding a greater voice in policy-making and, by their behavior, students are displaying an unwillingness to accept uncritically the dictates of their elders. In such a state of flux, the time is right for changed emphases and directions in the study of teachers and the learning process—not an inchoate search for a single, generalized pattern of qualities or behaviors that characterize good teachers or for sets of proved instructional techniques but for theory-building and indigenous studies which will transform institutional character.

STUDENT GAIN CRITERION

Every institution has purpose. Practices are introduced and maintained to further these purposes; people are selected to work within institutions in the expectation they will pursue organizational goals. In all cases, it is expected that something will result from institutional efforts.

In educational organizations, goals are broad and often nebulous. There has never been and never will be universal agreement on the goals of education. Each institution must, therefore, decide on its own objectives and translate them into operational terms, the goals then becoming the criterion variables upon which practices are adopted, policy is projected, and people selected (114).

Simply establishing objectives, however, and using them as criteria does not mean that selection of people, modes of organization, and policies can be undertaken. There are many types of criteria, some relating to process itself, others viewing product. Criteria may be such that the effort expended is assessed against them or that product is viewed through them. The ultimate criteria for evaluating an educational system might well be institutional ability to effect community transformation. Nevertheless, it would be an understatement to say it is difficult to create reliable measures of community change and to relate change to the efforts of a school.

Equally important—and certainly more easily measured—criteria are changes produced in students attending an educational institution. Even there, however, reliability is difficult to ascertain. The question of how students change under the impact of college is related to a broader question of the general conditions of personality change (109). Young people do change, personality develops, but relating those changes to life experiences in general (let alone to the effects of college) is a problem which psychological, sociological, and educational research has not begun to solve.

Development and testing of methodologies for assessing institutional effects must take place along several dimensions. Long-range community transformations and effects on students' personal development are serious questions to which people working with junior colleges can address themselves. But methodologies for assessing effects of instructors on even less far-reaching criteria must be undertaken as well. Because we have yet no reliable ways of assessing total college effects does not mean that we should fail to seek ways of assessing whether or not an institution, a department or a single instructor has had any effect in a particular direction on certain students.

The ultimate criteria for the junior college (as for any educational institution) are changes produced in its students and its community. The criteria may thus be viewed as *products* toward which the institution strives, rather than as *processes* in which it engages. Because of difficulties in obtaining reliable measures of long-range products, many writers in the field have settled on student gain toward specific objectives. These effects are variously called "student growth" or "student change" but they all involve measurement of change in student behaviors, actions, or abilities. Failure to attempt to measure student gain on the tangibles simply because the ultimate criteria do not lend themselves to reliable measurement, is futile and shortsighted.

THE ULTIMATE CRITERION

For purposes of study, then, measurable changes in students can be viewed as being ultimate criteria. These measurable changes may fall short of long-range effects, but they are much more closely related to those general transformations of personality and wisdom than are process or proximate criteria. Using student gain toward specific objectives as ultimate criteria has the additional value of being methodologically related to student change in general. As tools are developed to assess student gain in particular, measurement of broad-scale student change may be furthered. The two seem related. There is certainly face validity to the contention that measuring student gain toward specific objectives is more closely related to measuring student change in general than is assessing an institution or a single instructor on the basis of processes used or efforts expended.

Among investigators, the use of student gain on short-range objectives as a measure of teacher effectiveness is generally acknowledged as being more valid than the use of such criteria as, for example, teachers' effort expended or the various perceptions of observers. Orleans suggests:

> As the *ultimate* criteria of the effectiveness of the teacher's performance, we posit *the changes which take place in the behavior of pupils*. If the

> overall function of the educational process is to produce changes in the individuals, then the effectiveness of the teacher's performance must be measured by the extent to which it produces such changes (82:642).

The report of a committee of the American Education Research Association, set up to define problems in establishing criteria of teaching effectiveness, determined that:

> . . . a teacher's effect on the pupil's achievement of the immediate objectives of the given curriculum segment for which each teacher is respon.-sible is somewhat less ultimate

than the teachers effect on the student's total life, yet, it must be considered as essential in measurement. Similarly, Biddle sees teacher effectiveness as:

> . . . the ability of a teacher to produce agree upon educational effects in a given situation or context (9).

McKeachie, who has written as much about college teaching as anyone else on the contemporary scene, insists that "the ultimate criteria of effective teaching are changes in students in the direction of the goals of higher education" (72). And Anderson concludes:

> teacher evaluation experts are almost universally agreed that the measure of true effectiveness as a teacher is the change that is produced in the pupils taught by that teacher (2).

The list of educators who insist that student change must be considered as the ultimate criterion of teacher effectiveness could be extended. However, the issue cannot be settled merely by establishing validity of an arbitrary standard. Reliable measures for assessing teacher effects must still be produced. Efforts in finding them and problems in establishing them should not detract from the fact that assessing instructors on the basis of student change moves the entire issue of evaluation closer to the ultimate criteria of education.

MEASURING STUDENT GAIN

There are several problems in using student gain as a measure of teacher effectiveness. Most of those may be related to two broad issues: the kind of change (gain, learning) that shall be measured, and the way that student learning relates to the effects of an individual instructor.

What types of gain shall be measured? A case can be made for assessing general learning as measured by ability tests; but what of the instructor who teaches a particular subject area and makes no effort to affect students' general abilities? A determination to assess instructional effects only on the basis of instructors' specific objectives may resolve this issue. However, the concept of specifying measurable objectives is far from being universally accepted in the field ("The things *I* teach, you can't measure").

A related problem in selecting measures of student gain is in weighing the comparable worth of objectives. Students who are led to memorize data and show gains on measures of factual recall must be viewed as having learned in a different sphere from those showing gain on measures devised by an instructor who encourages them to analyze and to synthesize information. It is futile to consider using student gain on measures other than those which assess learning towards instructors' own objectives. If all instructors would specify clearly what they are trying to teach and what measures they intend to use to assess learning, arrangements could then be made to assess effects. If.

The other broad set of difficulties in assessing effect through student changes involves the problem of contamination. Are changes due solely to the influence of a particular instructor? Students may learn as a result of many things other than efforts expended by the instructor. Their learning is influenced by their general mental ability, past educational experiences, available instructional materials, influence of peers, socioeconomic background, types of extracurricular activities, quality of instruction in other areas of the curriculum, effects of mass media, and dozens of other variables (95).

In addition, any teacher in a classroom is himself an image whose effect may be altered by his students' perceptions of the total environment—perceptions often affected by prior experience. For example, a permissive teacher who has a positive effect on one student may have the opposite effect on another because his permissiveness reminds the student of a previous instructor who had exhibited those characteristics but failed the student! The suggestion, of course, is that the effective teacher is one who can alter his procedures to fit individual situations—a concept which was explored in studies reported earlier in this monograph.

A different type of problem relates to the commonly accepted definition of teaching as, "that which a teacher does." When the word "learning" is left out of the definition of teaching, resistance to evaluation on the basis of student gain must follow. Any scheme for assessing instructors demands the active cooperation of the faculty. If instructors feel they entered the profession to lecture, conduct discussions, etc.—that which a "teacher" typically does—assessment on the basis of what their students learn represents an alien dimension. The use of a student-gain assessment scheme may be threatening and thus, unfeasible.

If measures of student gain are to be related to an individual instructor's influence, the duration of time between pre- and post-teaching must be short. Therefore, other doubts may be raised as to the efficacy of this scheme. Some instructors have a genuine belief in their long-range effects. Their definition of learning is not "changed capability for, or tendency toward, acting in particular ways" but more like "something mysterious that may manifest itself at some unknown time." If, they reason, their effects may not be realized until years have passed, how rate them on any short-range basis?

There are other difficulties in using student gain as a criterion to assess instructors. Is the criterion related to the ability to cause learning in any and all situations or only in particular situations—for example, with certain types of teaching objectives or with certain types of students? There is clear import in that question because the teacher who fosters consistently high gain scores among "low-achieving" students is less frequently found in the profession than the teacher whose students could as likely learn without his influence. The analogy is that of the doctor who treats only well patients. Is he to be considered as valuable to society as one who takes *sick* people and makes them well? Again the purpose for which the assessment is conducted must be considered.

Another difficulty in using student gain as a criterion on which instructors are rated includes the fact that, to control for extraneous influences, the measure must be based solely on classroom activities. It must ignore advising, teachers' participation in extracurriculars, and other areas in which they come in contact with—and hence, influence—their students. It correspondingly fails to consider students' contacts with people other than the teacher. How isolate the variables?

As it relates to instructors' objectives, test validity also affects the feasibility of employing a student-gain criterion. The instructor who uses tests that assess his students' abilities to "analyze" is at a disadvantage compared to the instructor whose tests measure simple "recall." If external test builders come in and give tests that measure students' abilities at similar cognitive levels, although in different subject areas, the issue of relating tests to teacher objectives still remains (8). Dressel sums the problems posed in attempting to introduce a faculty assessment scheme based on student learning:

> The growth and development of students in regard to course objectives as measured by pre- and post-testing is one of the most attractive and logical means of evaluating teaching. Lack of adequate measures and the sheer work of collecting and analyzing the results limit the actual utility of the approach (32:12).

Nevertheless, if teachers *must* be judged—and the reasons for judging them vary almost as much as the schemes which are used—let them be evaluated on *effects* of their efforts, not on perceived worth of the efforts themselves. Observations—whether by colleagues, administrators, or committees representing both groups—are unreliable and invalid. The apparent difficulty in using student gain as a criterion should not encourage members of the profession to fall back exclusively on the use of proximate criteria.

No two people can agree on what the competent teacher is because, before beginning to judge competence, agreement on outcomes must be reached. Efforts in the junior colleges must be now devoted to defining types of desired effects and to measuring them. The attempt to define and assess the competence of an individual instructor or total faculty will suffer from sterility until the major task of identifying specific dimensions of the ultimate criterion has been undertaken.

STUDENT GAIN AND INSTRUCTIONAL SUPERVISION

chapter VI

NEW DIRECTIONS

Difficulties in using student gain as a major criterion on which to differentiate among instructors are not insurmountable. Certainly they are not so great that junior colleges must surrender to other considerably less potent schemes. It is possible to devise techniques for assessing teachers' abilities to cause student learning. Although less frequently found than other types of studies, a few have been reported in the literature.

One design, developed for use in Air Force schools, has become the prototype for studies that use student gain as a measure of instructors' effect (78). Instructors were selected on the basis of their teaching identical subject matter to selected students in similar classrooms, using the same training aids. Pretests were administered to all subjects and served as a basis for assessing learning. The chief results of these investigations suggested that:

> student gains can be reliably measured and that students' ratings of their instructors' teaching effectiveness and supervisors' ratings of instructors' verbal facility are correlated significantly with student gains. (78:IV).

And further:

> The high relationship . . . between ratings and rankings by fellow instructors and supervisors, together with the fact that these measures appear unrelated to student gains, suggest that *fellow instructors and supervisors judge instructor effectiveness on the basis of factors other than what students learn.* One of these factors appears to be the instructor's knowledge of subject matter. To obtain a completely adequate evaluation of an in-

structor it may be that a multiple criterion composed of supervisor ratings, student ratings, and student gains should be used (78:IV).

The Air Force studies were able to isolate teacher effects because all instructors worked toward identical learning objectives. Common examinations were administered. Both these procedures are necessary if teacher effects are to be identified without contamination.

A model currently being developed at U.C.L.A. separates from all possible and potential teacher activities just those which have direct effects on students. It also controls for teachers' overall knowledge of subject area and their knowledge about students with whom they are confronted. This design seeks evidence *only of teaching ability*—and in this case, even "teaching" is narrowed to include only teachers' abilities to prepare and present classroom lessons.

Using that design, Popham (92) gave sets of objectives and test items to two groups: secondary school teachers and graduate students with no prior teaching experience. Each instructor prepared and presented his lessons in any manner he considered appropriate. No significant differences in mean scores on the criterion examination were found between pupils taught by inexperienced teachers and those taught by the experienced group. It was concluded that because experienced teachers are not *set* to effect student learning toward specific objectives, they could do no better in a situation requiring such a task than could a group of people who had never taught before!

In another recently reported study, Justiz similarly prepared objectives and validated tests for a group of teachers in several high schools in the Los Angeles area (63). Again, unfamiliar objectives and test items were given to the teachers. Each instructor, working from common sets of objectives in an unfamiliar subject area with a specified amount of time for lesson preparation, was free to select content from a list of suggestions and to sequence his lesson in any way he chose. He was free to reveal objectives to the students, relate them in terms of their potential value, provide practice exercises, track students' practice work, and provide students with knowledge of results. In short, he could select any teaching technique.

For this study, classes were composed of students selected at random, their prior abilities or tendencies unknown to their teachers. Each instructor was thus required to present a lesson to unfamiliar students. The extent to which they learned was determined by using tests provided by the investigator. Correlations between rankings were significant at the .05 level of confidence.

This experiment demonstrated that when knowledge of subject area and of students is parcelled out and controlled along with extraneous influences, some teachers can consistently prepare and present lessons better than others. The variable tested had been narrowed down from the teacher as a "total person" to the instructor engaging in just those activities that conveyed the preparation and presentation of lessons. Accordingly, the evidence gained was that of pure classroom teaching. Instruction alone was the dependent variable. Each "good" teacher found his own best way of presenting his lessons. In the group of fourteen, there was no consistent teaching style but it was found that those who could teach one subject well could also teach the other. The design thus seemed to yield an index of generalized teaching ability.

PURPOSE

Designs which assess ability to prepare and present lessons only view one part of the total instructional process. However, the part they reliably measure is that which is supposedly being observed by a rater who sits in a classroom and watches the instructor. These procedures have potential for use in replacing the classroom observer who views instructor performance and *assumes* a connection between it and student learning. The assumption is somewhat less valid, and certainly less reliable, than the actual test.

Schemes that assess general teaching ability also can be used to select classroom specialists. If the profession of junior college teaching becomes specialized to the extent that one instructor is responsible for preparing objectives for a department, another for preparing test items, and so, then a measure of general teaching ability can be used to select specialists in the classroom. Whether or not such investigations are employed depends on the extent to which a particular junior college faculty is concerned with specialization and role differentiation. A design which can reliably select *effective* classroom instructors is a valuable aid to the *teaching* profession.

Although designs identifying teachers who can *teach* may be used to evaluate, rate, rank, or differentially reward practicing instructors, such measures seem short-sighted. Much more valuable to education would be studies relating people who exhibit particular teaching competencies to theories of human functioning. Who are the good teachers? Why are they so? Not only, "What classroom behaviors do they manifest?" but "What characteristics of personality do they possess?"

Various consortia of junior colleges are being founded around the country; they seem to have potential for joint efforts in the study of instructors and instructional processes. Whole studies can be replicated using different populations in different institutions; or each college might participate by taking a small part of a larger study coordinated by the consortium or by a university with which such a group might be affiliated. The point is that institutional efforts can be profitably employed in advancing the state of knowledge about people and about instruction. Those are much more pressing needs than continued unreliable, invalid "evaluations" of instructors.

INSTRUCTIONAL SUPERVISION

Supervision of instruction is rarely applied consistently in junior colleges. It is a spotty enterprise, too often subordinate to evaluation and confused with that wearisome endeavor. Many instructors are repelled by the idea that any facet of their instruction might be supervised by outsiders. They see no reason why it should be done and they view the entire process as being somehow like "public school." Yet the junior college is, by its own admission, a "teaching institution." Some type of instructional supervision seems warranted with the proviso, however, that it have deliberate purpose.

The junior college that is disinclined to participate in research on instructors and instructional processes may still abandon traditional methods of faculty evaluation and replace them with something of value. Most evaluation schemes are supposed to "improve instruction," although serious doubt exists as to their efficacy. But instructional improvement, a leading issue in junior college education (90), can be brought about through deliberate effort on the part of faculty and administrative leaders. A process of supervision with specific intent to cause particular changes in instructional practices can be the coordinating mechanism. It can foster communication between faculty and administration and place the college in a better position to answer the question, "Is anyone learning anything here?"

Communication between administration and faculty frequently suffers from misunderstanding and faulty perceptions. The supervisor who attempts to use evaluation as a means of communication often fails because evaluation is a judgmental process, hence, inevitably anxiety-provoking. The junior college dean or division chairman who must write faculty evaluations can make the evaluation process subordinate to general supervision to the benefit of both. He can bring the activity into a positive dimension by offering specific help to instructors rather than by standing in judgment of them. But instructors perceive supervisors' suggestions as being helpful only when the instructors themselves seek to cause student learning. Otherwise, all suggestions are perceived as being praise or blame—ends in themselves, and inadequate at best (127).

THE GOLDEN WEST PLAN

A supervisory scheme which seems to hold particular promise for enhancing communication and assessing instructional effects was introduced in 1967 at Golden West College (California). It was designed particularly to gain information so that resources could be properly allocated and so that institutional effects could be assessed. It was also intended as a curriculum-planning aid and as a way of leading instructors to specify their objectives—a worthy enterprise in its own right (22). Instructors voluntarily participate in this program.

The Golden West College scheme of instructional supervision begins at the initial employment interview. At that time instructors are informed about the college's move toward the definition of specific objectives in all courses and curriculums. They are encouraged to specify objectives in their own work. Because most instructors who are employed have not had previous training in writing objectives, arrangements are made for them to study the process. Help is provided by an experienced colleague, department chairmen, or outside expert.

A continuing series of scheduled interviews is arranged for both new and experienced instructors. At these meetings, the instructor, his division chairman and the dean of instruction review the instructor's objectives and the results he has obtained regarding student learning. Tests are examined and critiqued; objectives and media in use are reviewed. The situation is not one of threat but of aid. For example, an instructor who has sound objectives and testing devices but whose students are learning little may be given various forms of help, including suggestions as to various media and techniques which might be employed.

Meetings between department chairmen, dean, and individual instructors thus represent a continuing process of instructional supervision. Time allocated for the individual meetings is that which would ordinarily be used by the dean and the chairmen to visit classrooms and "observe." The Golden West College instructional supervision scheme replaces classroom observations and brings the process of supervision into meaningful perspective. It allows the administration to apportion resources in the form of aides and secretarial assistance; thus, in a sense, faculty members are rewarded for participating in the voluntary plan.

In reality, however, it is the process of instruction that is being rewarded. The program allows for in-service professional upgrading, for the introduction of new techniques of teaching, and for the development of experimental curriculums and instructional designs. It is actually supervision by the *objectives of instruction* rather than supervision of *instructors*. There is a marked distinction between the two (74).

Much learning theory and empirical data support the view that when there are clear statements of objectives, learning is enhanced. When objectives are reviewed *before* instruction, they can be brought closer to institutional purposes. When there is no agreement on (or understanding of) goals, observation of instructors in the classroom cannot be made within a valid frame of reference.

In the Golden West scheme, evaluation practices which depend upon singular variables independent of theory have given way to instructional supervision based on the extent to which students learn what their instructors attempt to teach. As the dean put it:

> It is felt that if one can evaluate what is happening to the learners, this is

more appropriate than evaluating teachers. The purpose of the school is for students to learn and this is the only outcome worth evaluating. (112).

NEXT STEPS

This monograph has taken the position that student gain toward specific learning objectives should be recognized as the ultimate criterion in assessing effects of teachers and teaching situations. It is a most defensible criterion variable because it relates directly to the acknowledged purposes of all educational endeavors. It is also a desirable dimension because it can help education as a whole move into sphere in which it can predict, manipulate, and accept accountability for its actions—in short, become a profession.

Teaching is acknowledged as being the main purpose of the junior college; study of instruction must, therefore, be undertaken. Exactly what is being learned in the junior college? By whom? Are curricular and instructional practices as effectual as they might be? For whom? What forms of student achievement should be accepted as evidence that learning has occurred. If students were provided with sets of specific objectives at the time of their entrance to the college, would their learning be enhanced? Would dropout be reduced?

Dimensions of people involved in the teaching-learning enterprise must be considered. How can particular types of instructors be prepared to cause learning? Should students be placed with instructors whose cognitive styles match their own? Or must "tracking" be done solely on the basis of prior achievement? Can everyone teach toward all types of learning objectives with equal facility? Or are some instructors better at causing recall, others at stimulating students to continue learning on their own? Why so? Is it the result of discernible actions or of personality characteristics which lend themselves only to indirect measurement?

Study of instructors and of instruction can merge in the junior college. Teacher evaluation along unspecified dimensions can and should give way to procedures and practices of much greater potential. In merging the two streams of study, the attempt to "describe" the act of teaching should be held separate because it has given rise to observational schemes from which no reliable inferences can be made.

The need now is to discover *who* can teach *whom*. Interactions of instructional situations must be identified. Prerequisite to the identification of effects is the specification of forms of learning to be accepted as evidence of attainment. Those classes of variables represent directions for potentially fruitful study in which junior colleges can profitably engage. To the extent they engage in those endeavors, junior colleges move toward enhancing the American educational enterprise.

BIBLIOGRAPHY

1. Adorno, T. W., et al. *The Authoritarian Personality*. New York: Harper, 1950.
2. Anderson, H. M. "A Study of Certain Criteria of Teaching Effectiveness." *Journal of Experimental Education 23:* 41-47 1954.
3. Arnstine, Donald G. "The Language and Values of Programmed Instruction." *The Educational Forum* 28:1964.
4. Astin, Alexander W., and Lee, Calvin B. T. "Current Practices in the Evaluation and Training of College Teachers." *The Educational Record* 47: 3; Summer 1966.
5. Barr, A. S. "Appraising Instructor Behavior." *The Appraisal of Teaching in Large Universities.* (Edited by W. J. McKeachie.) Ann Arbor: University of Michigan, 1959.
6. Barr, A. S., and Emans, Lester M. "What Qualities are Prerequisite to Success in Teaching?" *Nations Schools* 6: 60-64; September 1930.
7. Barr, A. S., et al. "Report of the Committee on the Criteria of Teacher Effectiveness." *Review of Educational Research* 22: 238-263; 1952.
8. ———. *Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness.* Madison, Wisconsin: Dembar Publications, Inc., 1961.
9. Biddle, Bruce J. "The Integration of Teacher Effectiveness Research." *Contemporary Research on Teacher Effectiveness.* (Edited by Biddle and Ellena.) New York: Holt, Rinehart and Winston, 1964.
10. Biddle, Bruce J., and Ellena, William J., editors. *Contemporary Research on Teacher Effectiveness.* New York: Holt, Rinehart and Winston, 1964.
11. Blocker, Clyde. "Are Our Faculties Competent?" *Junior College Journal* 36: 12-17; December 1965.
12. Bloom, Benjamin S., and others. *Taxonomy of Educational Objectives.* New York: David McKay Co., Inc., 1956.
13. Boyce, A. C. *Methods of Measuring Teachers' Efficiency.* Fourteenth Yearbook, National Society for the Study of Education. Chicago: University of Chicago Press, 1915.
14. Brawer, Florence B. *The Concept of Ego Strength and Its Measurement Through a Word Association Technique.* Doctor's Thesis. Los Angeles: University of California, 1967. Abstract: *Dissertation Abstracts* 28: 1699-A.
15. ———. *Personality Assessment of College and University Faculty: Implications for the Junior College.* Washington, D.C.: American Association of Junior Colleges, 1968.
16. Brawer, Florence, and Cohen, Arthur M. "Global and Sign Ap-

proaches to Rorschach Assessment of Beginning Teachers." *Journal of Projective Techniques and Personality Assessment* 30: 6; 1966.

17. Brown, James W., and Thornton, James W., Jr. *College Teaching*. New York: McGraw Hill Book Co., 1963.
18. Bruner, Jerome S. *Toward A Theory of Instruction*. Cambridge, Mass.: Harvard University Press, 1966.
19. Burkard, M. I. "Effectiveness of the MTAI in a Parochial School Setting." *Journal of Experimental Education* 33: 225-229; 1965.
20. Cattell, Raymond B. "Confirmation and Clarification of Primary Personality Factors." *Psychometrica* 12: 197-220; September 1947.
21. Coffman, W. E. "Determining Students' Concepts of Effective Teaching from Their Ratings of Instructors." *Journal of Educational Psychology* 45: 19; May 1954.
22. Cohen, Arthur M. "Defining Instructional Objectives." *Systems Approaches to Curriculum and Instruction in the Open-Door College*. (Edited by B. Lamar Johnson.) Junior College Leadership Program Occasional Report No. 9. Los Angeles: University of California, School of Education, 1967.
23. ———. "Developing Specialists In Learning." *Junior College Journal* 37: 21-23; September 1966.
24. Cohen, Arthur M., and Brawer, Florence B. "Adaptive Potential and First-Year Teaching Success." *Journal of Teacher Education* 43: 179-185; Summer 1967
25. ———. *Focus on Learning: Preparing Teachers for the Two Year College*. Los Angeles: University of California, Graduate School of Education, Junior College Leadership Program, 1967.
26. Cook, W. W.; Leeds, C. H.; and Callis, R. *The Minnesota Teacher Attitude Inventory*. New York: Psychological Corp., 1951.
27. Cooper, James G., and Lewis, R. B. "Quantitative Rorschach Factors in the Evaluation of Teacher Effectiveness." *Journal of Educational Research* 44: 703-707; 1951.
28. Dandes, Herbert M. "Psychological Health and Teaching Effectiveness." *The Journal of Teacher Education* 17: 301-306; Fall 1966.
29. Darley, J. G., and McNamara, W. J. *Minnesota Personality Scale*. New York: Psychological Corporation, 1941.
30. Davis, Hazel. "Evolution of Current Practices in Evaluating Teacher Competence." *Contemporary Research on Teacher Effectiveness*. (Edited by Bruce Biddle and William Ellena.) New York: Holt, Rinehart and Winston, 1964.
31. Dixon, Robert W., and Morse, William C. "The Prediction of

Teaching Performance Empathic Potential." *Journal of Teacher Education* 12: 322-329; September 1961.

32. Dressel, Paul. "The Current Status of Research on College and University Teaching." *The Appraisal of Teaching in Large Universities.* (Edited by W. J. McKeachie.) Ann Arbor: University of Michigan Press, 1959.
33. Dugan, Ruth R. "Personality and the Effective Teacher." *The Journal of Teacher Education* 12: 335-337; 1961.
34. Durflinger, G. W. "Personality Correlates of Success in Student Teaching." *Educational and Psychological Measurement* 23: 383-390; 1963.
35. Eiduson, Bernice T. *Scientists—Their Psychological World.* New York: Basic Books, Inc., 1962.
36. "Evaluating College Teaching." *Curriculum Reporter Supplement #1.* San Jose, Calif.: San Jose State College, December 1961.
37. Evans, Catharine, and McConnell, T. R. *Minnesota T-S-E Inventory, 1942-57.* Princeton, N.J.: Educational Testing Service, 1962.
38. Fielstra, Clarence. "Discriminative and Predictive Values of Ratings Given to UCLA Student Teachers on a Secondary School Level." *California Journal of Educational Research* 14: 11-18; 1963.
39. Fitch, Naomi. *Evaluation of Instructors in California Junior Colleges.* Berkeley: University of California, School of Education, Fall 1965. (Mimeo.)
40. Flanders, Ned A. "Some Relationships Among Teacher Influence, Pupil Attitudes and Achievement." *Contemporary Research on Teacher Effectiveness.* (Edited by Bruce Biddle and William Ellena.) New York: Holt, Rinehart, and Winston, 1964.
41. Gage, N. L., editor. *The Handbook of Research on Teaching.* Chicago: Rand McNally & Co., 1963.
42. Getzels, J. W., and Jackson, P. W. "The Teacher's Personality and Characteristics." *Handbook of Research on Teaching.* (Edited by N. L. Gage.) Chicago: Rand McNally & Co., 1963. pp. 506-582.
43. Gleazer, Edmond J., Jr. "Preparation of Junior College Teachers." *Educational Record* 48: Spring 1967.
44. Gough, H. G. *Manual for the California Psychological Inventory* (Rev., 1964). Palo Alto, Calif.: Consulting Psychologists Press, 1957.
45. Gough, H. G., and Heilbrum, A. B., Jr. *The Adjective Check List Manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1965.
46. Gough, H. G., and Pemberton, W. H. "Personality Characteris-

tics Related to Success in Practice Teaching." *Journal of Applied Psychology* 36: 307-309; 1952.

47. Gough, H. G.; Durflinger, Glenn W.; and Hill, Robert E., Jr. "Predicting Performance in Student Teaching from the California Psychological Inventory." *Journal Educational Psychology* 59: 119-127; 1968.
48. Guba, E. G., and Getzels, J. W. "Personality and Teacher Effectiveness: A Problem in Theoretical Research." *Journal of Educational Psychology* 46: 330-344; 1955.
49. Gustad, John W. "Policies and Practices in Faculty Evaluation." *The Educational Record* 42: 194-211; July 1961.
50. Guthrie, E. R., and Buxton, Claude E. *College Teaching: A Psychologist's View*. New York: Harcourt, Brace & Co., 1956
51. Guthrie, E. R. *The Evaluation of Teaching: A Progress Report*. Seattle: University of Washington Press, 1954.
52. Hahn, M. E., and Maclean, M. S. *Counseling Psychology*. New York: McGraw-Hill Book Co., Inc., 1955.
53. Harris, Chester W., and Liba, Marie R. *Encyclopedia of Educational Research*. (Third edition.) New York: MacMillan Co., 1960.
54. Harrower, M. R., and Steiner, M. E. *Large Scale Rorschach Techniques*. (Second edition.) Springfield, Ill.: Thomas, 1951.
55. Hathaway, S. R., and McKinley, J. C. *Minnesota Multiphasic Personality Inventory Manual*. (Revised 1951). New York: Psychological Corporation, 1943.
56. Heston, Joseph C. *Heston Personal Adjustment Inventory*. New York: Harcourt, Brace, and World, Inc., 1949.
57. Hill, R. E., Jr. "An Investigation of the California Psychological Inventory Empirically Keyed for Dichotomous Prediction of Student Teacher Excellence." *The Eighteenth Yearbook of the National Council on Measurements Used in Education*. (Edited by E. M. Huddleston.) Ames, Iowa: National Council on Measurements Used in Education, 1961.
58. "Dichotomous Prediction of Student Teaching Excellence Employing Selected CPI Scales." *Journal of Educational Research* 53: 349-351; 1960.
59. Howsan, Robert . *Who's A Good Teacher? Problems and Progress in Teacher Evaluation*. Burlingame, Calif.: Joint Committee on Personnel Procedures of the California School Board Association of the California Teachers Association, 1960.
60. Hoyt, D. P. *The Relationship Between College Grades and Adult Achievement: A Review of the Literature*. Iowa City, Iowa: American College Testing Program, 1965.
61. Jackson, Philip W. *The Teacher and The Machine*. Pittsburgh: University of Pittsburgh Press, 1968.

62. Joyce, B. R.; Lamb, H.; and Seiber, Joan. "Conceptual Systems and Information Processing: A Study of Teachers." *Journal of Educational Research* 59: 209-222; 1966.

63. Justiz, Thomas D. *A Method for Identifying the Effective Teacher.* Doctor's Thesis. Los Angeles: University of California, 1968.

64. Klopfer, W. G., and others. *Developments in the Rorschach Technique.* Yonkers-on-Hudson, N.Y.: World Book Co., 1954.

65. Knapp, Robert H. "Changing Functions of the College Professor." *The American College: A Psychological and Social Interpretation of the Higher Learning.* (Edited by N. Sanford.) New York: John Wiley and Sons, Inc., 1962. pp. 290-311.

66. Knapp, Robert H., and Goodrich, H. B. *The Collegiate Origins of American Scientists.* Chicago: University of Chicago Press, 1952.

67. Lamke, Tom Arthur. "Personality and Teaching Success." *Journal of Experimental Education* 20:217-259; 1951.

68. Laurits, James. "Thoughts on the Evaluation of Teaching." *Educational Horizons* 45: 87, 97; Spring 1967.

69. Leeds, C. H. "A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport." *Psychological Monographs* 64: 6; 1950.

70. Lonsdale, R. C. *An Evaluation of the Change in Educational Liberalism—Conservatism of Teachers-in-Service in a Situation Involving Off-Campus Courses and Workshops.* Doctor's Thesis. Syracuse, N.Y.: Syracuse University, 1952.

71. McKeachie, W. J. "Appraising Teaching Effectiveness." *The Appraisal of Teaching in Large Universities.* Ann Arbor: University of Michigan, 1959.

72. ———. "Procedures and Techniques of Teaching: A Survey of Experimental Studies." *The American College.* (Edited by N. Sanford.) New York: Wiley, 1966.

73. ———. "Psychology At Age 75: The Psychology Teachers Comes Into His Own." *American Psychologist* 23: 551-557; August 1968.

74. McNeil, John D. "Antidote to a School Scandal." *Educational Forum* 31: 69-77; November 1966.

75. Michaelis, J. U. "The Prediction of Student Teaching From Personality and Attitude Inventories." *University of California Publications in Education* 11: 415-481; 1954.

76. Minter, W., editor. *The Individual and the System.* Boulder, Colorado: Western Interstate Commission for Higher Education, November 1967.

77. Monroe, W. S. "Measuring Teacher Efficiency." *University of Illinois Bulletin.* Educational Research Circular No. 25. Urbana, Illinois: University of Illinois Press, 1924.

78. Morsh, Joseph E.; Burgess, George G.; and Smith, Paul N. *Student Achievement as a Measure of Instructor Effectiveness*. Project No. 7950. Task No. 77243. Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, n.d.

79. Murray, H. A., and others. *Explorations in Personality*. New York: Oxford University Press, 1938.

80. Myers, I. B., and Briggs, K. C. *Myers-Briggs Type Indicator*. Princeton, N.J.; Educational Testing Service, 1962.

81. National Education Association. Research Division. "The Teacher Looks At Personnel Administration." *Research Bulletin* 23: 93-147; December 1945.

82. Orleans, Jacob S., and others. "Some Preliminary Thoughts on the Criteria of Teacher Effectiveness." *Journal of Educational Research* 45: 641-648; May 1952.

83. Ort, Vergil K. "A Study of Some Techniques Used for Predicting the Success of Teachers." *Journal of Teacher Education* 15: 67-71; March 1964.

84. Overturf, C. L., and Price, Edwin C. *Student Rating of Faculty at St. Johns River Junior College*. Palatka, Florida: St. Johns River Junior College, Spring 1966. (Mimeo.)

85. Pace, C. Robert. "Five College Environments." *College Board Review* 41: 24-28; 1960.

86. ———. "Implications of Differences in Campus Atmosphere." *Personality Factors on the College Campus*. (Edited by R. L. Sutherland.) Austin: The Hogg Foundation for Mental Health, 1962.

87. Peck, Robert F. "Predicting Principals' Ratings of Teacher Performance From Personality Data." *Journal of Educational Psychology* 50: 70-74; 1959.

88. Peltier, Gary L. "Today's Salary Schedule." *The Educational Forum* 32: 417-421; May 1968.

89. Personal Orientation Inventory. *Manual for the Personal Orientation Inventory*. San Diego, Calif.: Educational and Industrial Testing Service, 1963.

90. Peterson, Basil. *Critical Problems and Needs of California Junior Colleges*. Sacramento: California Junior College Association, June 1965.

91. Philippi, H. A., and Childress, J. R. "A Letter to Teachers." *The Educational Forum* 32: 403-409; May 1968.

92. Popham, W. James. *Development of a Performance Test of Teaching Proficiency*. Final Report, U.S. Office of Education, Contract No. OE-6-10-254. Los Angeles: University of California, August 1967.

93. Rayder, Nicholas F. *College Student Rating of Instructors*. Ad-

dress given at Michigan State University, 1967. East Lansing, Mich.: Michigan State University, 1967.

94. Reavis, W. C., and Cooper, D. H. *Evaluation of Teacher Merit in City School Systems.* Supplementary Educational Monograph No. 59. Chicago: University of Chicago, 1945.

95. Remmers, H. H. "On Students' Perceptions of Teachers' Effectiveness" *The Appraisal of Teaching in Large Universities.* Michigan, 1959. pp. 19-20.

96. Riley, John W., Jr.; Ryan, Bruce F.; and Lifshitz, Marcia. *The Student Looks at His Teacher; An Inquiry Into the Implications of Student Ratings at the College Level.* New Brunswick, N.J.: Rutgers University Press, 1950.

97. Roe, Anne. "Early Determinants of Vocational Choice." *Journal of Counseling Psychology* 4: 212-217; Fall 1957.

98. ———. *The Psychology of Occupations.* New York: Wiley, 1956.

99. Rokeach, M. *The Open and Closed Mind: Investigations Into the Nature of Belief Systems and Personality Systems.* New York: Basic Books, 1960.

100. Rorschach, H. *Psychodiagnostics.* (Fifth edition.) Berne: Hans Huber, 1951.

101. Rosencranz, Howie A., and Biddle, Bruce J. "The Role Approach to Teacher Competence." *Contemporary Research on Teacher Effectiveness.* (Edited by Bruce Biddle and William Ellena.) New York: Holt, Rinehart, & Winston, 1964.

102. Rosenzweig, S. "The Picture Association Method in the Application in the Study of Reactions to Frustrations." *Journal of Personality* 14: 3-23; 1945.

103. Rostker, L. E. "The Measurement of Teaching Ability: Study #1." *General Experimental Education* 14: 6-51; 1945.

104. Ryans, D. G. *Characteristics of Teachers.* Washington, D.C.: American Council on Education, 1960.

105. ———. "Investigation of Teacher Characteristics." *Educational Record* 34:371-396; October 1953.

106. ———. *Measurement and Prediction of Teacher Effectiveness.* Proceedings of Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1959. pp. 52-67.

107. ———. "Research on Teacher Behavior in the Context of the Teacher Characteristics Study." *Contemporary Research on Teacher Effectiveness.* (Edited by Bruce Biddle and William Ellena.) New York: Holt, Rinehart, and Winston, 1964.

108. Sandgren, D. L., and Schmidt, L. G. "Does Practice Teaching Change Attitudes Toward Teaching?" *Journal of Educational Research* 49: 673-680; 1956.

109. Sanford, Nevitt. "Higher Education as a Social Problem."

American College. (Edited by Nevitt Sanford.) New York: John Wiley & Sons, 1962.

110. Schroder, H. M.; Driver, M.; and Streufert, S. *Information Processing in Individuals in Groups.* New York: Holt, Rine-Hart, and Winston, 1967.
111. Seagoe, May V. "Prediction of In-Service Success in Teaching." *Journal of Educational Research* 39: 658-663; 1946.
112. Shawl, William F. *Project Proposal.* Huntington Beach, Calif.: Golden West College, 1968. (Mimeo.)
113. Simpson, Ray H., and Seidman, Jerome M. *Student Evaluation of Teaching and Learning.* Washington, D.C.: The American Association of Colleges for Teacher Education, 1966.
114. Sorenson, A. Garth. "A New Role in Education: The Evaluator." *UCLA Evaluation Comment.* Los Angeles: University of California, Center for the Study of Evaluation of Instructional Programs, January 1968.
115. Sprinthall, N. A.; Whiteley, J. M.; and Mosher, R. L. "A Study of Teacher Effectiveness." *Journal of Teacher Education* 17: 19; Spring 1966.
116. Stecklein, John E. "Colleges and Universities—Programs." *Encyclopedia of Educational Research.* New York: The Macmillan Co., 1960.
117. Stern, George G.; Stein, Marvin J.; and Bloom, Benjamin S. *Methods in Personality Assessment.* (Revised edition.) Glencoe, Illinois: The Free Press, 1963.
118. Super, D. E., and Crites, J. O. *Appraising Vocational Fitness.* New York: Harper and Row, 1962.
119. Symonds, Percival M. "Teaching As A Function of the Teacher's Personality." *Journal of Teacher Education* 5:79-83; March 1954.
120. Symonds, Percival M., and Dudek, Stephanie. "Use of the Rorschach in the Diagnosis of Teacher Effectiveness." *Journal of Projective Techniques* 20: 227-234; 1956.
121. Tanner, William C., Jr. "Personality Bases in Teacher Selection." *Phi Delta Kappan* 35: 271-277; April 1954.
122. Teilhet, Raul. "A Letter From Teachers." *The Educational Forum* 32: 409-415; May 1968.
123. Travers, Robert M. W., and Wallen, Norman E. "Analysis and Investigation of Teaching Methods." *Handbook of Research on Teaching.* (Edited by N. L. Gage.) Chicago: Rand-McNally & Co., 1963. pp. 448-505.
124. Trent, James W., and Medsker, Leland L. *Beyond High School.* San Francisco, Calif.: Jossey-Bass, Inc., 1968.
125. Tyler, F. T. "The Prediction of Student Teaching Success From Personality Inventories." *University of California Publications in Education* 11: 233-313; 1954.

126. Veldman, D. J., and Peck, R. F. "Student-Teacher Characteristics. From the Pupil's Viewpoint." *Journal of Educational Psychology* 54:346-355; 1963.

127. Wittrock, Merlin C. "Set Applied to Student Teaching." *Journal of Educational Psychology* 53: 175-180; 1962.

128. Zamansky, H. S., and Goldman, A. E. "A Comparison of Two Methods of Analyzing Rorschach Data in Assessing Therapeutic Change." *Journal of Projective Techniques* 24: 75-83; March 1960.